UNIWERSYTET
WARSZAWSKI

Novel data processing methods for randomly sampled multi-dimensional NMR experiments

JAN STANEK

Warsaw, September 2014

Jan Stanek, *Novel data processing methods for randomly sampled multi-dimensional NMR experiments*
Tytuł w języku polskim: *Nowe metody przetwarzania losowo próbkowanych wielowymiarowych eksperymentów NMR*

Ph. D. thesis under supervision of
Prof. Wiktor Koźmiński
Faculty of Chemistry
Biological and Chemical Research Centre
University of Warsaw
Pasteur 1, 02-093 Warsaw

The results presented in this thesis were obtained with the generous financial support of the European Comission's 7th FrameWork projects:

East-NMR

Bio-NMR

This thesis was prepared using LaTeX 2_ε word processing sytem.

Contents

Preface

It is constantly surprising that resolving potential of numerous multidimensional NMR experiments has not been recognized for years. With the gaining popularity and dissemination of non-classical signal sampling, many of those methods have become useful, creating increased interest in signal processing strategies. The motivation for this work was to improve quality of spectral data obtained from NUS experiments.

The thesis is divided into four chapters, of which the first two are primarily literature based, and the remaining ones present the results of own research.

In the first chapter the major challenges in biomolecular NMR are presented, as well as the most impressive and relevant advances that appeared in the field in the last 15-20 years. It is shown that owing to both methodological and hardware improvement, NMR is increasingly more sensitive and more effective at addressing key problems in structural biology. However, the inherent problem of signal crowding in NMR spectra of large macromolecules is still to be solved.

Chapter two introduces the issue of sampling limitations that, until recently, prevented from effective use of high-dimensional NMR methods. The possible solutions employing non-uniform sampling and other methods are summarized. In the following, advantages and deficiencies of various NUS data processing methods are discussed. The origins of the algorithm proposed in this thesis and its relationship to alternative methods are presented.

Chapter three discusses the algorithm design and implementation. Additionally, comparisons to selected alternative processing methods are shown.

Finally, chapter four focuses on the applications of proposed processing method. Certainly, it is particular relevance to reveal and promote new applications since it can increase the popularity and practical relevance of NUS and developed processing algorithms. Most examples in this chapter demonstrate applications to NUS four- and five-dimensional spectra, with far less emphasis on three-dimensional methods. This is due to the fact that the limited resolution enhancement possible for the latter renders it quite inattractive. In contrast, more spectacular results are obtained for 4D+ experiments, which were, up to recently, on the edge or behind the limits of practical application.

List of author's publications (chronological order)

Original papers

1. **Jan Stanek**, Wiktor Koźmiński, *Iterative algorithm of discrete Fourier transform for processing randomly sampled NMR data sets*, Journal of Biomol. NMR 47 (2010) 65–77

2. **Jan Stanek**, Rafał Augustyniak, Wiktor Koźminski, *Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm*, Journal of Magnetic Resonance 214 (2012) 91–102

3. Wolfgang Bermel, Ivano Bertini, Isabella C. Felli, Leonardo Gonnelli, Wiktor Koźmiński, Alessandro Piai, Roberta Pierattelli, **Jan Stanek**, *Speeding up sequence specific assignment of IDPs*, Journal of Biomolecular NMR 53 (2012) 293–301

4. Dmytro Lozhko, **Jan Stanek**, Krzysztof Kazimierczuk, Anna Zawadzka-Kazimierczuk, Wiktor Koźminski, Igor Zhukov, Alexander Kornelyuk, *^{1}H, ^{13}C, and ^{15}N chemical shifts assignments for human endothelial monocyte-activating polypeptide EMAP II*, Biomolecular NMR Assignments 7 (2013) 25–29

5. **Jan Stanek**, Saurabh Saxena, Leonhard Geist, Robert Konrat, Wiktor Koźmiński, *Probing of Local Backbone Geometries in Intrinsically Disordered Proteins by Cross-Correlated NMR Relaxation*, Angewandte Chemie Int. Ed. 52 (2013) 5404–06

6. **Jan Stanek**, Michał Nowakowski, Saurabh Saxena, Katarzyna Ruszczyńska-Bartnik, Andrzej Ejchart, Wiktor Koźmiński, *Selective diagonal-free ^{13}C,^{13}C-edited aliphatic-aromatic NOESY experiment with non-uniform sampling*, Journal of Biomolecular NMR 56 (2013) 217–226

7. **Jan Stanek**, Peter Podbevšek, Wiktor Koźmiński, Janez Plavec, Mirko Cevec *4D Non-Uniformly Sampled C,C-NOESY Experiment for Sequential Assignment of ^{13}C,^{15}N-Labeled RNAs*, Journal of Biomolecular NMR 57 (2013) 1–9

8. Z. Orbán-Németh, M. A. Henen, L. Geist, S. Żerko, S. Saxena, **J. Stanek**, W. Koźmiński, F. Propst, R. Konrat, *Backbone and partial side chain assignment of the microtubule binding domain of the MAP1B light chain*, Biomol. NMR Assignm. 8 (2014) 123–127

9. Saurabh Saxena, **Jan Stanek**, Mirko Cevec, Janez Plavec, Wiktor Koźmiński, *C4'/H4' selective, non-uniformly sampled 4D HC(P)CH experiment for sequential assignments of ^{13}C-labeled RNAs*, Journal of Biomolecular NMR, *accepted*

Reviews

10. Krzysztof Kazimierczuk, **Jan Stanek**, Anna Zawadzka-Kazimierczuk, Wiktor Koźmiński, *Random sampling in multidimensional NMR spectroscopy*, Progress in Nuclear Magnetic Resonance Spectroscopy 57 (2010) 420–434

11. Krzysztof Kazimierczuk, Maria Misiak, **Jan Stanek**, Anna Zawadzka-Kazimierczuk, Wiktor Koźmiński, *Generalized Fourier transform for non-uniform sampled data*, Topics in Current Chemistry 316 (2012) 79–124

12. Krzysztof Kazimierczuk, **Jan Stanek**, Anna Zawadzka-Kazimierczuk, Wiktor Koźmiński, *High-dimensional NMR spectra for structural studies of biomolecules*, ChemPhysChem 14 (2013) 3015–3025

Acknowledgements

I wish to express my gratitude to my supervisor, *Wiktor*, for educating, inspiring and motivating me. I am also indebted to my colleagues and co-workers:

Ania
Marysia
Mateusz
Michał
Karolina
Kasia
Krzysztof Kazimierczuk
Krzysztof Kosiński
Saurabh
Szymon
Wojtek

– for their help, creative discussions and friendly atmosphere created in our group. I also wish to acknowledge my Polish and foreign collaborators from ENS Paris (*Rafał, Fabien*), Vienna University (*Leo, Robert*), IBB Warsaw (*Andrzej, Kasia, Igor*) KI Ljubljana (*Peter, Mirko, Primož, Janez*), CERM Firenze (*Alessandro, Isabella*) and IMB Kiev (*Dmytro*) – for their substantial contribution to the work showed in this thesis.

Finally, I am particularly grateful to my beloved *wife* for her support, understanding and infinite patience.

Abbreviations

The following common literature abbreviations are used throughout the thesis:

APSY – automated projection spectroscopy
BEST – band-selective excitation short transient (experiment)
BMRB – Biological Magnetic Resonance Bank
CCR – cross-correlated relaxation
COSY – correlation spectroscopy
CR – cross-relaxation
CRINEPT – cross relaxation-enhanced polarization transfer
CRIPT – cross relaxation induced polarization transfer
CS – compressed sensing *or* chemical shift
CSA – chemical shift anisotropy
CT – constant time (evolution)
DD – dipole-dipole *or* dipolar
DFT – discrete Fourier transformation
DQ – double-quantum
HD – hyperdimensional
HMQC – heteronuclear multiple-quantum correlation
HSQC – heteronuclear single-quantum correlation
IDP – intrinsically disordered protein
INEPT – insensitive nuclei enhanced by polarization transfer
IRLS – iterative re-weighted least squares (algorithm)
IST – iterative soft thresholding
iFT, $(\mathrm{FT})^{-1}$ – inverse Fourier transformation
MAS – magical angle spinning
MaxEnt – maximum entropy method/reconstruction
MDD – multidimensional decomposition
MQ – multiple-quantum
NOE, NOESY – nuclear Overhauser effect (SpectroscopY)
nuFT – non-uniform Fourier transformation
NUS – non-uniform sampling *or* non-uniformly sampled
PDB – protein data bank
PFG – pulsed field gradient
PRE – paramagnetic relaxation enhancement
PSF – point spread function
RDC – residual dipolar coupling
SCT – semi-constant time (evolution)
SE – sensitivity enhancement (Rance-Kay method)
SIFT – spectroscopy by integration of frequency and time-domain information
SSA – signal separation algorithm
TOCSY – total correlation spectroscopy
TROSY – transverse relaxation-optimized spectroscopy
ZQ – zero-quantum
z.a. FT – zero-augmented Fourier transformation (a. k. a. nuFT)

Three-letter aminoacid names and one-letter codes of ribonucleotides are used.

Abstract

The topic of this dissertation is a new algorithm for processing of sparsely sampled data sets from multidimensional nuclear magnetic resonance (NMR) experiments. NMR remains one of the major experimental technique for studying biological macromolecules. However, increasing size of investigated objects poses a challenge for NMR due to rapidly decreasing sensitivity and increasing signal crowding. The first chapter focuses on recent advances in sensitivity enhancements and summarises a few solutions for resolution of spectral overlap. Subsequently, one describes the crucial and limiting problem of signal sampling in multidimensional NMR, which, up to recently, has impeded the widespread use of high-dimensional NMR methods. Major fast acquisition and non-uniform sampling (NUS) approaches are presented. The particular emphasis was put on detailed discussion of competetive approaches to processing of data from NUS experiments. In chapter 3 the new iterative algorithm is proposed for artefact suppression in high-resolution NMR spectra. The detailed description of its design and implementation is given, and followed by comparison with selected processing methods. The efficacy of the algorithm is demonstrated on model synthetic and experimental data. The last chapter of the thesis shows various applications of the proposed method to existing and new four- and five-dimensional NMR experiments. The algorithm is proven most beneficial in challenging applications including spectra for assignment of side-chain resonances in protein and nucleic acids, NOESY spectra for structural analysis, and cross-correlated relaxation measurements for proteins.

Streszczenie

Niniejsza praca jest poświęcona nowej metodzie przetwarzania danych pochodzących z oszczędnie próbkowanych wielowymiarowych eksperymentów jądrowego rezonansu magnetycznego (ang. *Nuclear Magnetic Resonance*, NMR). Technika ta jest, obok krystalografii rentgenowskiej, główną eksperymentalną metodą badawczą pozwalającą na określenie struktury i dynamiki makromolekuł o znaczeniu biologicznym. Jednakże NMR napotyka dwie istotne przeszkody w odniesieniu do dużych biomolekuł, a mianowicie gwałtownie pogarszającą się czułość oraz krytyczne zatłoczenie sygnałów w widmach. W rozdziale pierwszym przedstawiono ostatnie osiągnięcia w poprawie czułości technik NMR oraz rozwiązania służące podniesieniu rozdzielczości widm. Następnie opisano kluczowy problem próbkowania wielowymiarowych sygnałów NMR, który do niedawna uniemożliwiał wykorzystanie pełnego potencjału tych technik do rozdzielenia sygnałów. Omówiono pokrótce współczesne podejścia do szybkiej akwizycji i oszczędnego próbkowania sygnałów NMR (ang. *non-uniform sampling*, NUS). Szczególny nacisk położono na porównanie i dyskusję wad i zalet stosowanych obecnie metod przetwarzania sygnałów niejednorodnie próbkowanych. W rozdziale 3-cim opisano nowy iteracyjny algorytm oparty o transformację Fouriera, usuwający artefakty oszczędnego próbkowania w wysokorozdzielczych widmach NMR. Szczegółowo omówiono schemat algorytmu oraz jego programową implementację. Rozdział uzupełnia porównanie wyników algorytmu oraz wybranych metod przetwarzania na wysymulowanych oraz modelowych danych eksperymentalnych. W ostatnim rozdziale pracy zademonstrowano użyteczność nowej metody do literaturowych oraz nowych cztero- i pięciowymiarowych eksperymentów NMR. Wśród proponowanych zastosowań wymienić można widma do przypisania sygnałów w łańcuchach bocznych aminokwasów (w białkach) i pierścieniach rybozy (w kwasach rybonukleinowych), widma NOESY służące określeniu struktury trójwymiarowej biomolekuł, oraz pomiary szybkości relaksacji skorelowanej w łańcuchach głównych białek.

Chapter 1

Challenges and recent advances in biomolecular NMR spectroscopy

1.1. Wealth of NMR observables

The major premise for aiding structural biology with nuclear magnetic resonance spectroscopy is (*i*) its ability to investigate biomolecules in their nearly native environment, or (*ii*) the case if they fail to form well behaving crystals. The particular utility of NMR manifests in atomic resolution of acquired information, irrespective whether dynamics, interactions or structure are studied. In the following, a compact list of parameters derived from various NMR techniques is provided, as well as interpretations that can be assigned. This set is by no means complete, and lacks a detailed explanation of obtained quantities, however, sheds some light on the versatility and relevance of NMR in the biomolecular research.

Chemical shifts ($\delta_H, \delta_C, \delta_N, \delta_P$) These basic parameters can be easily accessed and measured with high precision. However, detailed quantification is hampered by the multiplicity of factors influencing CSs. Nevertheless, they are clear indicators of aggregation state (monomer, dimer, etc.), and presence of stable tertiary structure. CSs, C^{α}-$^{13}C\beta$ differences in particular, inform on propensity to form secondary structure elements (Wishart et al 1992; Wishart and Sykes 1994). The backbone chemical shifts also correlate satisfactorily with dihedral angles (φ, ψ). These dependencies are valid for triples of aminoacids, i.e. the residue and its neighbours. TALOS is one of the programs that analyses CSs in this respect (Shen et al 2009). An important application is mapping of chemical shift perturbations (CSP) which occur upon ligand binding or formation of multimeric proteins from subunits. This fairly straightforward method alone often suffices to detemine domain interfaces in protein complexes.

Nuclear Overhauser Effect Significant NOE reflects the spatial proximity of nuclear spins, usually protons. As explained in sec. 1.3.2, the volume of NOE cross-peak scales with inverse sixth power of interatomic distance. To roughly derive the distances the relative cross-peak volume (w. r. t. the auto-correlation one) is compared to a cross-peak in a reference moiety with a fixed geometry, such as geminal protons in CH_2 groups or aromatic rings:

$$r_{12} = r_{\mathrm{ref}} \left(\frac{\sigma_{\mathrm{ref}}}{\sigma_{12}} \right)^{1/6} = r_{\mathrm{ref}} \left(\frac{V_{\mathrm{ref}}}{V_{12}} \right)^{1/6}$$

NOE remains the major source of structural information in NMR of stably folded proteins (Wüthrich 1986). However, it is unlikely to provide contacts between spins distant more than

$\sim$6 Å. In the multimeric proteins, intermolecular NOEs are of particular interest to determine binding interface.

Scalar couplings Three-bond (dihedral) coupling constants, e.g. $^3J_{\mathrm{H}\alpha,\mathrm{H^N}}$, correlate well with backbone dihedral angles φ, ψ. Experimental parametrization of so-called Karplus curves are required (Karplus 1963). Frequently, ambiguous angles are derived from a single J value. Also one- and two-bond heteronuclear couplings, e.g. $^2J_{N,C\alpha}$, proved to be useful in structural analysis of proteins (Wirmer and Schwalbe 2002).

Residual dipolar couplings These quantities can be measured in weakly aligning (anisotropic) media (Tjandra and Bax 1997). Alternatively, weak magnetic-field induced alignment can be observed at high fields due to the anisotropy of molecular magnetic susceptibility, especially for RNAs. RDC for a pair of spins may be expressed as:

$$D_{ij}(\theta,\varphi) = -\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij}^3}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta cos(2\varphi)\right) \tag{1.1}$$

where $A_a = \frac{1}{2}A_{zz}$ and $A_r = \frac{1}{3}(A_{xx} - A_{yy})$ are *axial* and *rhombic* components of alignment tensor $\mathbf{A}$, and θ, φ are polar angles defining orientation of internuclear vector $\vec{r_{ij}}$ w. r. t. principal axes of $\mathbf{A}$. Thus, a single RDC indicates possible vector orientations on a cone, and measurement of different kinds of RDS (e.g. $\mathrm{H^N}$-^{15}N, and ^{15}N-C') or in multiple media is advised (Ding and Gronenborn 2004).

J-couplings and RDSs play important role in structure determination of perdeuterated proteins, where limited number of $\mathrm{H^N}$-$\mathrm{H^N}$ NOEs hinders precise structure calculations.

Paramagnetic constraints Due to presence of a paramagnetic centre, either a metal ion or an intentionally introduced spin label (tag), relaxation of proximal spins is enhanced. The contribution due to dipolar interaction with an unpaired electron (PRE) can be expressed as:

$$R_{2p} = \frac{9}{32}\frac{\gamma_I^2 g_S^2 \hbar^2}{r^6}\left(4J(0) + 3J(\omega_I)\right) \tag{1.2}$$

where r is a nuclear-electron distance and J is the spectral density function. Also, pseudo-contact shifts (PCSs) are observed due to fast averaging of chemical shifts of unequally populated states with opposite electron spin orientation. PREs and PCSs are valuable long-distance constraints: usually the nuclei in the sphere of $\sim$20Å are affected (Battiste and Wagner 2000; Liang et al 2006).

Auto-correlated relaxation In general, T_1, T_2 carry information on internal motions at ps-ns timescales. Routinely, the measurements also include the $\{^1\mathrm{H}\}$-^{15}N NOE experiment to quantify $\mathrm{H^N}$-^{15}N cross-relaxation rate σ_{NH}, or NOE enhancement factor, at steady-state conditions. Auto- and cross-relaxation rates generally depend on the values of spectral correlation functions $J(\omega)$ at different frequencies, depending on the involved relaxation mechanisms. E.g., dipolar $\mathrm{H^N}$-^{15}N relaxation of in-phase N^+ coherence contributes as

$$R_2 \propto 4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)$$

The respective rate for cross-relaxation rate (NOE) is given by eq. 1.11. The overall goal of the relaxation experiments is to sample spectral density function at possibly many frequencies, which requires multiple B_0 fields.

The model for internal motions has to be introduced for interpretation of spectral density function values. In the simple model-free approach a complete separation of the internal and global motions is assumed, resulting in expression

$$J(\omega;\tau_c,\tau_e,S)=\frac{2}{5}\left(\frac{S^2\tau_c}{1+\omega^2\tau_c^2}+\frac{(1-S^2)\tau}{1+\omega^2\tau^2}\right) \tag{1.3}$$

where $\tau^{-1}=\tau_c^{-1}+\tau_e^{-1}$, τ_c and τ_e are correlation time constants (ns) for global and internal motions (ps-ns), respectively (Lipari and Szabo 1982). S^2 is the Lipari-Szabo order parameter, which can be interpreted as "local rigidity" ($0\leq S\leq 1$). If $J(\omega)$ is sampled at multiple frequencies, more advanced models of internal motions can be employed.

Cross-correlated relaxation CCR rates quantify the correlation between different relaxation mechanisms. Dipolar or CSA-induced fields may correlate if their fluctuations occur due to common molecular motions. The extent of correlation indicates (*i*) rigidity of the particular structure element, and (*ii*) mutual orientations of the respective magnetic (DD or CSA) fields. Usually, CCR rates are employed to derive backbone dihedral angles (see par. 4.3) or to detect concerted motions (Vögeli and Yao 2009).

Relaxation dispersion Measurement of rotating-frame relaxation $R_{1\rho}$ can provide data about chemical exchange if $k_{\mathrm{ex}}\approx\omega_e$, where ω_e is the rotating-frame nutation frequency. Relaxation dispersion is measured by varying spin-locking B_1 field (ω_1, usually up to 5 kHz is possible) or carrier offset. In the case of highly skewed populations in a two-site exchange ($p_1\gg p_2$) the measured rates are given by

$$R_{1\rho}=R_1\cos^2\theta+R_2\sin^2\theta+\frac{p_1p_2\Delta\omega^2k_{\mathrm{ex}}}{k_{\mathrm{ex}}^2+\omega_1^2+\Omega_2^2}\sin^2\theta \tag{1.4}$$

where R_1 are R_2 refer to the relaxation of the dominant state, θ is the angle between B_0 and effective (off-resonance) field, and k_{ex} denotes the exchange rate. In this specific, but important case, $R_{1\rho}$ depends on the resonance offset of the "minor" state, Ω_2. Also, the formula is valid for slow, intermediate and fast exchange. Populations, exchange rate and chemical shifts can be determined by fitting $R_{1\rho}$ dispersion curve, even if the "minor" state cannot be directly observed. $R_{1\rho}$ measurements are increasingly popular in the studies of protein *excited states* or short-living *folding intermediates*.

1.2. Increasing complexity of NMR spectra of biomolecules

1.2.1. Spectral crowding: impact of size, similarity and order

Atomic resolution of derived information is at the same time the blessing and curse of biomolecular NMR. Proteins and nucleic acids are linear polymers which implies the limited deviations of specific chemical shifts[1]. Thus, unavoidably, increasing size of the biomolecule leads to resonance crowding. Obviously, resonance overlap might greatly complicate site-specific interpretation of NMR spectra.

The major source of chem. shift differences are electronic effects, i.e. directly associated with type and number of chemical bonds. In the case of backbone ^{15}N and ^{13}C′ spins one

[1]Current histograms of ^{1}H, ^{13}C, and ^{15}N chemical shifts in proteins and nucleic acids can be accessed from Biological Magnetic Resonance Bank (http://bmrb.wisc.edu)

observes a great scatter since they depend on types of the residue and its neighbour (20 × 20 combinations in proteins). However, electronic effects on side-chain resonances (e.g. $^{13}C\beta$) are characteristic for a given type of aminoacid. As a result, signal dispersion in ^{13}C-HSQC spectra is quite limited (as compared to ^{15}N-HSQC), and mainly due to the presence of stable tertiary structure. Noteworthy, fast molecular motions such as aromatic ring flipping and methyl group rotation lead to dynamic averaging of particular chemical shifts, thus further reducing their diversity.

Fig. 1.1 shows aromatic region of ^{13}C-HSQCs of three globular proteins of greatly increasing complexity: S100A1 (2×93 aa), LipaseA (181 aa) and MBP (370 aa). Despite small

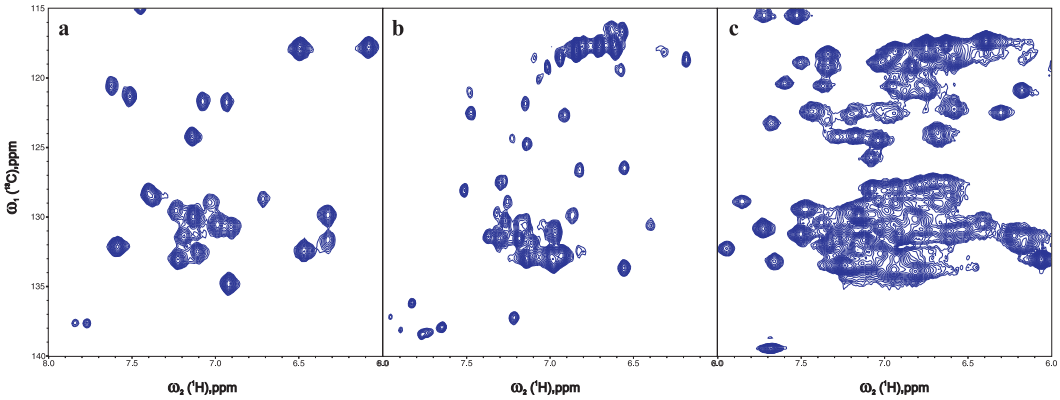


Fig. 1.1 Aromatic region of ^{13}C-HSQC spectra of S100A1 (a), LipaseA (b) and Maltose Binding (c) proteins .

correlation of aromatic ^{1}H and ^{13}C shifts, the spectrum becomes heavily crowded for large proteins (MBP), making it problematic to resolve peaks in the most populated regions. The situation becomes critical in the case of aliphatic resonances (see Fig. 1.2). ^{13}C-HSQC spectrum is reasonably resolved only for relatively small proteins such as ubiquitin (Fig. 1.2a). For larger proteins (e.g. LipaseA), only methyl region retains acceptable resolution. However, in fully protonated large proteins (MBP) increased ^{1}H linewidths and rapidly growing number of resonances limits the utility of this region as well (Fig. 1.2c).

Apart from the size, also intrinsic chemical diversity of a biomolecule may affect the signal crowding. For example, primary sequences of many IDPs are highly repetitive, causing increased abundance of specific aminoacids (Pro, Lys, Asp, Asn, Glu, Gln). Another example are nucleic acids which comprise only four kinds of nucleotides (A, G, C, T/U). Fig. 1.3 shows the peak dispersion in H1'/C1' and aromatic regions of ^{13}C-HSQC of moderately sized 34-nt

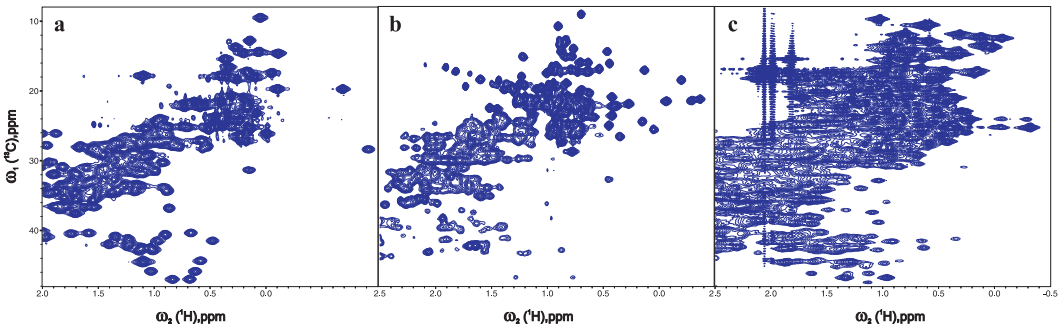


Fig. 1.2 Subregion of aliphatic ^{13}C-HSQC spectra of ubiquitin (a), LipaseA (b) and Maltose Binding (c) proteins showing methyl and some of methylene resonances.

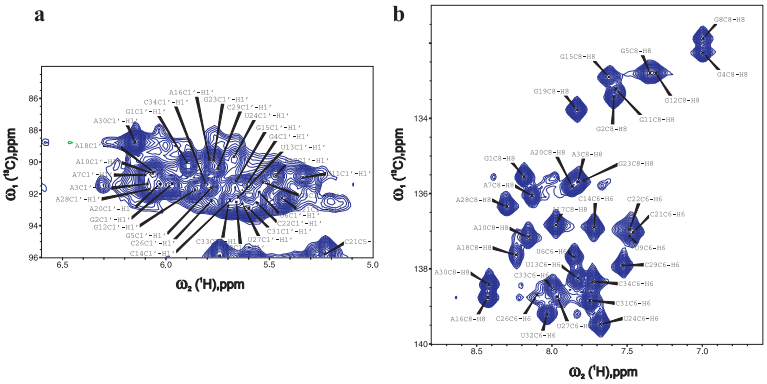


Fig. 1.3 Ribose H1'/C1' (*a*) and aromatic H6/C6 and H8/C8 (*b*) regions of ^{13}C-HSQC spectra of 34-nucleotide RNA *LCS1co* (Cevec et al 2010).

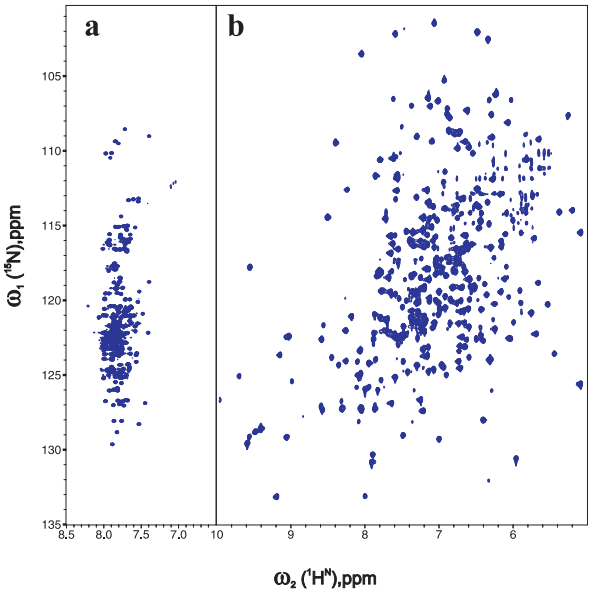


Fig. 1.4 Comparison of ^{15}N-HSQC spectra of typical large proteins featuring a labile and stable tertiary structure, respectively: Chicken BASP1 (270 aa) (*a*) and Maltose Binding Protein (370 aa) (*b*).

non-coding RNA. Due to chemical similarity of consecutive ribose rings (of only *one type*), H1'/C1' region suffers from considerable peak overlap, and it is even more pronounced in other regions (H2', H3', H4', H5' – not shown). On the contrary, the aromatic region (Fig. 1.3b) exhibits resolution similar to proteins due to a comparable number of ring types (3 and 4).

Finally, the effect of local magnetic fields should be considered. Such anisotropic fields may be induced by currents in aromatic rings or in side-chain and backbone carboxyl groups. As shown in Fig. 1.4b, a stable tertiary structure may lead to excellent signal dispersion even in large proteins (MBP). However, in the case of IDPs, the structural motifs are formed transiently, and the role of local fields diminishes. Particularly affected are H^N (and side-chain ^{1}Hs) shifts, which effectively average towards random-coil values, as demonstrated in Fig. 1.4a. A similar situation occurs in non-coding RNAs, where the absence of base-stacking also results in decreased dispersion of resonances.

1.2.2. Relaxation, line widths and sensitivity

The factor that has been silently ignored in the previous section is the impact of increasing molecular mass onto resonance natural linewidths. In the slow tumbling limit, the dipolar ^{1}H transverse relaxation is governed by eq.

$$\frac{1}{T_2} = \frac{1}{20} d^2 J(0) \tag{1.5}$$

where $d = \hbar\gamma_H^2 r^{-3}$ and r is the inter-proton distance.

Assuming isotropic rotational diffusion and a rigid spherical rotor, the spectral density function J is given by

$$J(\omega) = \frac{2}{5}\frac{\tau_C}{1+\omega^2\tau_C^2} \tag{1.6}$$

According to Stokes-Einstein equation, the rotational correlation time τ_C is given by:

$$\tau_C = \frac{4\pi\eta}{3kT}R^3 \tag{1.7}$$

where η, T, k stand for solvent viscosity, temperature and Boltzmann constant, and R is the hydrodynamic radius of a globular protein. The above equations lead to the following proportionality:

$$\Delta\omega = \frac{1}{\pi T_2} \propto \tau_C \propto R^3 \propto M \propto n \tag{1.8}$$

i.e. the natural ^{1}H linewidths $\Delta\omega$ scale linearly with molecular mass M, and thereby also with a number of ^{1}H spins n (Rule and Kevin 2006). This unfavourable relationship contributes into peak overlap in NMR spectra of large molecules. Decreased lifetimes of coherences not only degrade resolution, but also exponentially decrease sensitivity of pulse sequence schemes, where fixed delays are required for magnetization transfer.

Additionally, the longitudinal relaxation is adversely affected by increasing molecular mass:

$$\frac{1}{T_1} = \frac{6}{20}d^2\left(\frac{4}{5}\frac{\tau_C}{1+\omega_H^2\tau_C^2} + \frac{4}{5}\frac{4\tau_C}{1+4\omega_H^2\tau_C^2}\right) \tag{1.9}$$

The terms on the right-hand side decrease for sufficiently large τ_C, effectively T_1 increases with τ_C, and mass. The longer recovery time affects the steady-state magnetization at the optimal pulsing condition (see eq. 2.19), decreasing sensitivity as

$$\mathrm{S/N} \propto \frac{1}{\sqrt{T_1}} \tag{1.10}$$

In conclusion, large biomolecules pose a real challenge for solution-state NMR due to linearly increasing crowding of spectra, linewidths, and rapidly degrading sensitivity.

1.3. Solutions (1): Isotope labelling and coherence transfer

Various isotope labelling schemes represent one of possible remedies to tackle both complexity and insensitivity of biomolecular NMR. The simplification of spectra may be accomplished by restricting the number of magnetically active spins (see par. 1.3.1) or by resolving spectra into additional heteronuclear dimensions using coherence transfer (par. 1.3.2 and 1.3.3).

1.3.1. Heteronuclear labelling and deuteration

The introduction of *uniform* ^{15}N,^{13}C-labelling in the late 80's (McIntosh and Dahlquist 1990) has triggered a rapid development of NMR experiments and fostered new biomolecular applications. This emphasises the relevance of biosynthetic and cell-free methods of selective isotope labelling for advances in NMR. The progress in this field is still continued, providing solutions to complexity and insensitivity of NMR of large biomolecules.

In the mid 90's, the applicability of NMR has been extended beyond the limit of ~20 kDa mass by the use of fractional (LeMaster 1994) and perdeuteration of non-exchangeable protons (Venters et al 1995). Almost complete replacement of ^{1}H spins by deuterons boosted

sensitivity of NMR methods (see par. 1.4.2). However, very limited structural information is available in the absence of side-chain protons. Unfortunately, studies employing randomly fractionally protonated samples encounter difficulties associated with lower spin density and multiplicity of isotopomers. Therefore, alternatives based on selective reprotonation have been sought as a more practical solution to structure determination of large proteins.

The particular attention has been attracted by specific methyl protonation since (*i*) having typically 35-45% abundance in the primary sequence and (*ii*) preferable distribution of CH_3 groups in the hydrophobic core of the proteins, they represent excellent probes of structure and interactions. Additionally, (*iii*) 3-fold degeneracy of methyl ^{1}Hs and presence of (*iv*) the methyl-TROSY effect considerably increase sensitivity of methyl detected experiments.

An early approach relied on use of protonated pyruvate as the only source of ^{13}C, leading to selectively protonated (otherwise deuterated) methyl sites of Ala, Val, Leu and Ile (γ_2 only), however, with a substantial residual protonation of Hβ (Rosen et al 1996). More recently, a more cost-effective approach employing [^{1}H,^{13}C]-D-glucose and 100% D_2O was proposed to label all methyl sites at deuteration level between 50 and 80% (Otten et al 2010). Clearly, samples require isotopomer filtration and the method is not applicable to proteins of mass exceeding ∼40 kDa.

The most popular protocols employ deuterated glucose, 100% D_2O, and a suitable a.a. precursor(s) to selectively reprotonate aminoacids according to their biosynthetic pathways. The most prominent methods were developed in L. Kay's laboratory. In the initial protocol of Gardner and Kay (1997) (3,3-$^2H_2$) 2-ketobutyrate was added to selectively label δ_1 group of Ile, while supplementing the minimal media with Val yielded desired protonation pattern of Val and Leu. The subsequent optimization proposed by Goto et al (1999) relied on the use of [3-^{2}H] α-ketoisovalerate to label Val and Leu with higher yields, thus at lower cost. Importantly, ^{13}C may be introduced to all sites or restricted to methyl groups. For larger proteins such as 723 a.a. MSG (Tugarinov et al 2003), 2-keto-3-methyl-d_3-3-d_1-butyrate is used to introduce only one protonated methyl group into Leu and Val, and thus restrict intraresidue ^{1}H-^{1}H interactions (Tugarinov and Kay 2004). Nevertheless, the commercially available precursor is a racemic mixture of pro-*S* and pro-*R* compounds, leading to presence of all signals in the fingerprint ^{13}C-HSQC. These patterns are commonly abbreviated as "ILV"-labelling as a short-hand notation for [^{1}H,^{13}C]-Ile,Leu,Val(δ_1)-[^{2}H,^{15}N,^{12}C]-labelling.

The recent advances in the field aim at even greater simplification of spectra. E.g., the selective and full protonation of only one prochiral (either pro-*S* or pro-*R*) groups of Val and Leu (Gans et al 2010) reduces twice the complexity of ^{13}C-HSQC. Other important developments are methods for regiospecific labelling of Ile-γ_2 (Ruschak et al 2010; Ayala et al 2012), Met-ϵ (Fischer et al 2007), Ala-β (Isaacson et al 2007; Ayala et al 2009) and Thr-γ (Velyvis et al 2012; Sinha et al 2011) methyl sites, using appropriate precursors. If these labelling patters are insufficient to identify (assign) CH_3 groups, one can resort to single-point mutagenesis as demonstrated by Amero et al (2011).

Well-isolated methyl groups are crucial for structure determination as they provide sensitive and quantitative NOE, unaffected by spin diffusion even at long mixing time (Tugarinov et al 2005a). An alternative, yet quite expensive, approach is based on prochiral synthetic aminoacids, which form the so-called stereo-array isotope labeling (SAIL) (Kainosho et al 2006). This method yields both stereospecific assignments in all aminoacids, and stereospecific NOE contacts, which significantly improve the precision of NMR-derived 3D structure.

The complexity of resonance assignment can be also solved by complete but aminoacid-selective protonation or combinatorial ^{13}C and ^{15}N labelling (Löhr et al 2012; Parker et al 2004). Another important application of isotope labelling is identification of interdomain (or intermolecular) NOEs by either mixed [^{13}C,^{15}N]-/[^{12}C,^{14}N]- or [^{1}H]-/[^{2}H]-labelling (Nudel-

man et al 2011) of individual subunits. Isotope labelling can also solve some notorious problems encountered in NMR spectroscopy. For example, the alternate ^{13}C-^{12}C-labelling proposed by LeMaster and Kushlan (1996) removes homonuclear ^{13}C-^{13}C couplings, thus allowing high resolution in ^{13}C dimensions.

1.3.2. Coherence transfer

The key role of coherence transfer is to enable chemical shift labelling of *correlated* nuclei in order to enhance spectral resolution. Physical phenomenon behind *correlation* is a relatively strong scalar coupling between spins, usually directly bonded, e.g. H$^{\mathrm{N}}$-^{15}N, H$^{\alpha}$-C$^{\alpha}$, C$^{\alpha}$-C'.

Coherence transfer may also be employed to gain information on coupling topology, e.g. dipolar-coupled networks of interacting ^{1}H-^{1}H pairs (NOE), ^{13}C-^{13}C scalar coupled spins in side-chains of aromatic and aliphatic residues in proteins and nucleic acids (TOCSY), or ^{2}J(C$^{\alpha}$N)- and ^{1}J(C$^{\alpha}$N)-coupled spins which usually establish sequential connectivities between backbone resonances. The reasonable value and uniformity of *heteronuclear* scalar coupling constants (cf. Fig. 1.5) enabled development of numerous triple-resonance techniques and greatly simplified resonance assignment of proteins (Ikura et al 1990; Kay et al 1990b).

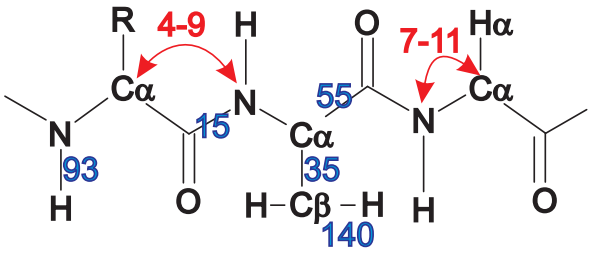


Fig. 1.5 Essential one- and two-bond scalar coupling constants employed for coherence transfer in double-labelled proteins. Adapted from Sattler et al (1999).

For brevity, well-recognized methods based on weak scalar couplings (i.e. INEPT, HMQC, DEPT) or isotropic mixing of strong coupling conditions (TOCSY) will not be discussed here. Instead, in the following the emphasis will be put on cross-relaxation induced (NOE) transfer, and more recently proposed cross-correlated relaxation induced (CRIPT) and spin-state to spin-state selective polarization transfers (ST2-PT). These transfer types are of particular relevance for large molecules.

Dipolar cross-relaxation induced Dipolar interaction between two proximal spins causes fluctuating magnetic field as a molecule tumbles in solution. Not only results it in efficient auto-relaxation of coherences in I-S system, but also induces cross-relaxation between I_z and S_z populations (and only between them). It means that T_1-relaxation of two dipolar interacting spins is not independent, and the corresponding cross-relaxation rate depends on difference of zero- and double-quantum transition rates (probabilites):

$$\Gamma_{\mathrm{IS}}^{DD} = W_2 - W_0 = \frac{6d_{00}}{4} J(\omega_I + \omega_S) - \frac{d_{00}}{4} J(\omega_I - \omega_S) \tag{1.11}$$

where $d_{00} = (\mu_0/4\pi)^2 \hbar^2 \gamma_I^2 \gamma_S^2 r_{IS}^{-6}$. The substitution of J from eq. 1.6 into 1.11 yields

$$\Gamma_{\mathrm{IS}}^{DD} = \frac{d_{00}}{10} \tau_C \left(\frac{6}{1 + (\omega_I + \omega_S)^2 \tau_C^2} - \frac{1}{1 + (\omega_I - \omega_S)^2 \tau_C^2} \right) \tag{1.12}$$

Since dipolar interaction is most pronounced for high-γ nuclei, the cross-relaxation effects are usually considered for homonuclear spin system (I=S=^{1}H, $\omega_S \approx \omega_I \approx \omega_H$). In this case the

formula (1.12) simplifies to

$$\sigma_{\mathrm{HH}}^{\mathrm{NOE}} \equiv \Gamma_{\mathrm{HH}}^{DD} = \frac{d_{00}}{10}\tau_C \left(\frac{6}{1+4\omega_H^2\tau_C^2} - 1 \right)$$

Biomolecules at high B_0 field are in the the spin diffusion limit ($\omega_0\tau_c \gg 1$), thus the first term (double-quantum transition rate) can be neglected and the cross-relaxation rate reduces to:

$$\sigma_{\mathrm{HH}}^{\mathrm{NOE}} = -\frac{\hbar^2\mu_0^2\gamma_H^4\tau_c}{160\pi^2 r_{\mathrm{HH}}^6} \tag{1.13}$$

Dipolar cross-relaxation manifests in *Nuclear Overhauser Effect* (NOE). After initial perturbation of the system, both ^{1}H spins are stored as $\sigma_A = \sigma_{1A} + \sigma_{2A} = a\mathrm{H}_{1z} + b\mathrm{H}_{2z}$ and left for a certain *NOE mixing time*. The density matrix evolves into $\sigma_B = (ra + \epsilon b)\mathrm{H}_{1z} + (rb + \epsilon a)\mathrm{H}_{2z}$, where factors r and ϵ are due to auto- and cross-relaxation, respectively. In effect, NOE triggers through-space coherence transfer between protons 1 and 2.

One can infer from the eq. 1.13 that the NOE build-up rate increases with correlation time, and consequently, with the size of a biomolecule. The possiblity to obtain large NOE during a reasonably short mixing time (typically 50-200 ms) makes NOE spectroscopy particularly useful for large systems (Xu et al 2005, 2006).

Induced by relaxation interference (CRINEPT) Relaxation theory predicts that the cross-correlation of dipole-dipole and CSA relaxation can lead to conversion of in-phase into anti-phase magnetization (Goldman 1984):

$$\mathrm{H}_x \xrightarrow{DD/CSA} 2\mathrm{H}_x\mathrm{N}_z$$

The anti-phase term arises as:

$$\langle 2\mathrm{H}_x\mathrm{N}_z\rangle(\tau) = \sinh(\Gamma_C\tau)\exp(-R_{\mathrm{H}}\tau)\langle \mathrm{H}_x\rangle(0) \tag{1.14}$$

where $\Gamma_C = \frac{4}{15}(\gamma_{\mathrm{H}}B_0\Delta\sigma_{\mathrm{H}})(\hbar\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}r_{\mathrm{NH}}^{-3})\tau_C$ is the cross-correlated relaxation rate, σ_{H} is the CSA of ^{1}H, and R_{H} is the auto-relaxation rate of ^{1}H. The possibility of useful coherence transfer using CSA/DD cross-correlation was demonstrated by Dalvit (1992), and simple spin echo can be used for this purpose (see Fig. 1.6b). For comparison, INEPT (Fig. 1.6a) evolves magnetization as follows (Morris and Freeman 1979):

$$\langle 2\mathrm{H}_y\mathrm{N}_z\rangle(\tau) = \sin(\pi J_{\mathrm{NH}}\tau)\exp(-R_{\mathrm{H}}\tau)\langle \mathrm{H}_x\rangle(0) \tag{1.15}$$

One crucial difference should be noted here: since R_{H} is proportional to correlation time τ_C, the efficiency of INEPT degrades rapidly for slowly tumbling molecules. In contrast, R_{H} and Γ_C in eq. 1.14, both proportional to τ_C, appear always in product with the transfer time τ. Consequently, transfer efficiency of CRIPT can be always optimized to the same value by adjusting τ. As shown by Riek et al (1999), despite rather small H^{N} CSA of $\Delta\sigma \approx 15$ ppm, CRIPT overcomes INEPT for $^1\mathrm{H}\rightarrow{}^{15}\mathrm{N}$ transfer if correlation times exceeds $\sim$100 ns. Optimal CSA/DD cross-correlation, thus maximal CRIPT transfer, is obtained at B_0 corresponding to ^{1}H frequency of 1.1 GHz. In practice, biomolecules of correlation times larger than 300 ns are rarely studied. In the intermediate regime of 50-300 ns both CRIPT and INEPT may contribute significantly to observable signal. Riek et al (1999) demonstrated that both pathways can be utilized using a surprisingly simple scheme shown in Fig. 1.6c. The cross-correlated relaxation-enhanced polarization transfer (CRINEPT) has the efficiency

$$A(\tau) = \sqrt{\sinh(\Gamma_C\tau)^2 + \sin(\pi J_{\mathrm{NH}}\tau)^2}\exp(-R_{\mathrm{H}}\tau) \tag{1.16}$$

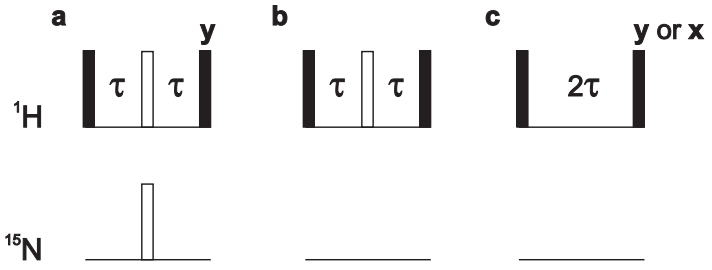


Fig. 1.6 Comparison of INEPT (a), CRIPT (b), and CRINEPT (c) blocks. In the last case the ^{1}H chemical shift evolution has to be compensated using additional pulse sequence elements as described by Riek et al (1999).

thus clearly taking benefit of both transfer types. Riek et al (1999) also demonstrated that CRINEPT can be incorporated into TROSY to fully exploit line-narrowing effects of CSA/DD interference in evolved ^{15}N and ^{1}H dimensions. ^{15}N-CRINEPT-TROSY was applied to 110 kDa protein of $\tau_C \approx 70$ ns. It was also noted that ^{15}N-HMQC inherently employs two CRINEPT transfers. If used without ^{15}N decoupling during acquisition, HMQC may actually outperform CRINEPT-TROSY for intermediate τ_C. However, TROSY effect in ^{15}N evolution (t_1) cannot be exploited as it is possible in CRINEPT-TROSY.

Spin-state selective coherence transfer As outlined in par. 1.4.2, selective coherence transfer between particular spin-state coherences (e.g. $\mathrm{N}^+\mathrm{H}^\beta$) may be desired to exploit benefits of their favourable relaxation properties. The scheme which accomplishes the transfers:

$$\mathrm{I}^\beta S^\mp \xrightarrow{\pi I_y S^\alpha} I^\mp S^\mp \xrightarrow{\pi I^\alpha S_y} \mathrm{I}^\mp \mathrm{S}^\beta \tag{1.17}$$

is presented in Fig. 1.7, and consists of two spin-state-selective coherence transfer modules (S^3CT) introduced by Sørensen et al (1997). The first S^3CT block effects in a S^α-selective π pulse on spin I ($\pi I_y S^\alpha$), while the second one is a H^α-selective π pulse on spin S ($\pi I^\alpha S_y$). Simultaneously, undesired anti-TROSY coherence is transferred as follows

$$\mathrm{I}^\alpha S^\mp \xrightarrow{\pi I_y S^\alpha} I^\pm S^\mp \xrightarrow{\pi I^\alpha S_y} \mathrm{I}^\pm \mathrm{S}^\alpha \tag{1.18}$$

As demonstrated by Pervushin et al (1998b) and Weigelt (1998), the separation of TROSY and anti-TROSY paths can be obtained using PFGs: gradient echo retains path 1.17 while suppressing 1.18. By convention, during direct detection only I^- is detectable. Therefore, gradient-enhanced *single-transition-to-single-transition polarization transfer* (ST2-PT) selects $\mathrm{I}^\beta S^- \rightarrow \mathrm{I}^-\mathrm{S}^\beta$ path. To obtain phase-sensitive detection in ^{15}N dimension, phases ϕ_1

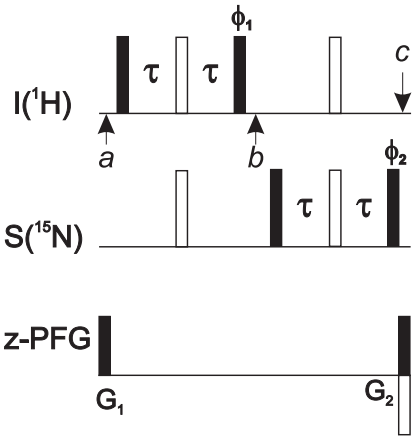


Fig. 1.7 Pulse scheme for ST2-PT. Both phases ϕ_1 and ϕ_2 are y and $-y$ for coherence transfers as in eq. 1.17 and 1.19, respectively. Different phase conventions for Bruker and Varian/Agilent spectrometers necessitate inversion of ϕ_1 and ϕ_2.

and ψ_2 in both S^3CT blocks have to be inverted. The modified scheme can be summarized as follows:

$$I^{\beta}S^{\mp} \xrightarrow{\pi I_y S^{\beta}} I^{\pm}S^{\mp} \xrightarrow{\pi I^{\beta} S_y} I^{\pm}S^{\beta} \tag{1.19}$$

In this case, TROSY path is retained by anti-echo gradient selection. Meissner et al (1998) proposed the three-gradient scheme, where the additional PFG pulse at time b in Fig. 1.7 distinguishes TROSY and anti-TROSY pathways, which exist as DQ and ZQ coherences in this point, respectively. Noteworthy is also the alternative transfer scheme of Nietlispach (2005), which avoids intermediate storage of fast relaxing multiple-quantum coherences and thus features higher sensitivity. In the case of insignificant DD/CSA interference of I spin, e.g. for aromatic ^{1}H, the half-TROSY scheme is employed (Meissner and Sørensen 1999).

1.3.3. Increased dimensionality

Additional evolution periods can greatly resolve peak overlap present in the spectra of lower dimensionality. The prerequisite for this approach is the possibility to trigger effective coherence transfer, thus it also implies the use of heteronuclear labelling. Contrary to homonuclear methods, ^{13}C or ^{15}N-resolved spectra contain the same number of peaks (e.g. 3D HNCO), or greater by a factor of two (3D HNCA) or three (3D (H)N(CA)NH), than their lower-dimensional counterparts (2D ^{15}N-HSQC). Effective resolution enhancement leans on the extent of correlation between chemical shifts in the added and original dimensions. For example, the following extensions substantially improve peak dispersion: 2D ^{13}C-HSQC → 3D $H^{\alpha}C^{\alpha}CO$, 2D ^{15}N-HSQC → 3D HNCO, or 2D [^{1}H, ^{1}H]-NOESY → 3D ^{15}N-edited NOESY-HSQC. In contrast, 3D CACON usually exhibits similar dispersion to 2D CON if applied for IDPs due to high correlation of C^{α} and C'.

In general, at least four difficulties arise that may impede effective use of higher-dimensional spectra. Usually, additional coherence transfer steps are required, which leads to signal attenuation. In some cases, however, the new sequences utilize the already present transfer periods, e.g. 3D (H)CCH-TOCSY → 4D HCCH-TOCSY, 4D HabCab(CO)NH → 5D HabCabCONH, 3D (H)N(CACO)NH → 5D HN(CA)CONH.

The second difficulty is the need for extended phase cycle to ensure unambiguous coherence transfer. For practical reasons, its length is usually compromised in favour of the number of sampling points (increments). However, the use of PFGs can partially compensate for deficiencies of short phase cycles. The third issue are sampling requirements which, at high field spectrometers, limit the attainable resolution already in 3D spectra, and until recently, prevented the use of higher-dimensional experiments. The problem and possible solutions are discussed in sec. 2.1.

Finally, it should be emphasised that increased dimensionality always implies concomitant loss of sensitivity per unit of time. Let me first consider the constant (non-modulated) signal with noise. The FT spectrum of this signal will show a single resonance at the centre ($\omega = 0$), cf. Fig. 1.8. Let the signal to noise in a single scan to be x, and peak intensity equal to 1. Accumulation of the second scan gives $S(\omega = 0) = 2$ and noise r.m.s.d. intensity $\sqrt{2}x$. Let me know consider the additional signal modulation (for higher dimensionality). The phase sensitive detection requires two signal modulations: cosine and sine. The real-valued cosine spectrum[2] contains two lines of halved intensity $S(\pm\Omega) = \frac{1}{2}$, and noise x, while in the imaginary part of sine spectrum $S(\pm\Omega) = \pm\frac{1}{2}$. The co-addition of these spectra $(\mathcal{R}(\mathcal{FT}(cos(t)) + i\mathcal{I}(\mathcal{FT}(\sin(t))$ yields the real part of the phase-sensitive spectrum. Obtained

[2]Both imaginary part of cosine modulation and real part of sine modulation vanish due to antisymmetry of Fourier integral

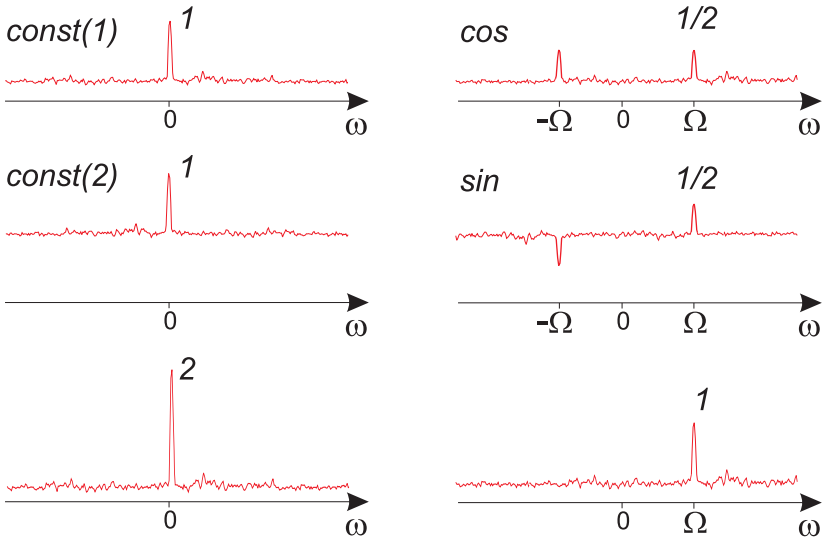


Fig. 1.8 The comparison of Fourier spectra of two unmodulated (constant) signals (*top and middle left* panels), and cosine and sine modulated ones (*top and middle right* panels). Resulting spectra are shown in *bottom* panels with respective relative intensities of the resonance peaks.

signal intensity is $S(\Omega) = 1$, i.e. 2-fold smaller than without modulation. The noise is accumulated in the same time in both cases, thus its intensity is identical and amounts to $\sqrt{2}x$. Therefore, signal-to-noise per unit of time decreases by a factor of 2 with every added dimension.

The inherent sensitivity loss can be mitigated by a factor of $\sqrt{2}$ (at best) by the use of so-called sensitivity-enhanced schemes (Palmer III et al 1991; Kay et al 1992). In this approach, both orthogonal spin operators (S_x and S_y) are transferred for detection using independent paths. The processing requires additional linear combination of *echo* and *anti-echo* phase-modulated signals.

1.4. Solutions (2): sensitivity and resolution enhancement

1.4.1. Dynamic nuclear polarization (DNP)

Low gyromagnetic ratio of nuclear spins results in low equilibrium polarization, and thus is the major reason for insensitivity of NMR. In contrast, electron spin polarization is greater by a factor of $\gamma_e/\gamma_n \approx 660$ at the same B_0 field, and can in favourable conditions be transferred to nuclear spins. The underlying process in liquids, involving dipolar coupling between nuclear and electron spins, was theoretically predicted by Overhauser (1953) and soon experimentally confirmed. The phenomenon is termed *dynamic nuclear polarization* (DNP).

In liquids, the transfer of magnetization from a free radical (such as nitroxides) to a nuclear spin is mediated by dipolar cross-relaxation, similar to internuclear NOE effect described in par. 1.3.2. A crucial cross-relaxation rate is related to difference in the "forbidden" zero- and double-quantum transitions: $\sigma_{e-n} = W_2 - W_0$. Under continuous microwave (MW) irradiation of allowed electron spin transitions, the steady-state nuclear polarization enhancement is given by the formula (Hausser and Stehlik 1968):

$$\epsilon = \frac{I_z - I_0}{I_0} = -\xi\, f\, s\, \frac{\gamma_e}{\gamma_n} \tag{1.20}$$

The coupling factor ξ determines the efficiency of the cross-relaxation compared to nuclear auto-relaxation ρ_n:

$$\xi = \frac{\sigma_{e-n}}{\rho_n} = \frac{W_2 - W_0}{W_0 + 2W_n + W_2} \tag{1.21}$$

where W_n is the allowed nuclear single quantum transition. The leakage factor f describes the ratio of nuclear relaxation associated with the presence of electron, compared to other relaxation mechanisms W_{ext}:

$$f = \frac{\rho_n}{\rho_n + W_{\text{ext}}} \tag{1.22}$$

This factor tends to 1 with typical concentrations (mM) of paramagnetic species. The saturation factor s accounts for the efficiency of MW pumping, and can be expressed in the simplest case as

$$s = \frac{S_0 - S_z}{S_z} = \frac{\eta}{1+\eta} \tag{1.23}$$

where $\eta = \gamma_e^2 B_1^2 T_{1e} T_{2e}$ depends on microwave transverse field component B_1 and electron longitudinal and transverse relaxation times T_{1e} and T_{2e}. Unfortunately, its hard to saturate electron transitions in nitroxide radicals due to fast relaxation (lifetimes of order of ns), hyperfine splitting to 3 lines, and high power requirements which can be detrimental for biological samples. The coupling factor ξ is responsible for low efficiency of DNP at high fields. For purely dipolar interaction, $\xi = 0.5$ for $B_0 = 0$ and decreases proportionally to B_0^2 or $B_0^{3/2}$, depending on translational diffusion of radical and target molecule.

In conclusion, simplistic and advanced theoretical models predict very small DNP efficiency at fields stronger than 5T. Despite different enhancement mechanisms, similar conclusions were drawn for solid state. Therefore, until recently DNP was not seriously considered for sensitivity enhancement of high-field NMR.

The situation changed dramatically with the invention of gyrotrons in 1990's, which enabled stable high-power MW irradiation at frequencies above 100 GHz. Currently, up to 525 GHz gyrotrons are used which limit ^{1}H DNP NMR to 800 MHz (installed in Lyon and Utrecht). Pioneering works of R. Griffin at MIT showed that in ssNMR at low temperatures (∼90 K) DNP enhancements as large as 300 can be obtained (Maly et al 2008). Nitroxide biradicals such as TOTAPOL were shown to boost DNP efficiency in solids (Hu et al 2004), and their optimization is an active field of research (e.g. in Lyon laboratory).

The breakthrough in liquid-state NMR was the invention of *dissolution* DNP (Ardenkjær-Larsen et al 2003). In this method, the sample is frozen (at $T \approx 1.6$ K) and polarized at low field in a separate magnet (3.4 T) within hours. Subsequently, it is rapidly dissolved and diluted with "hot" solvent, and transferred to NMR magnet for detection. Despite tremendous singal enhancements exceeding 10 000, the experiment cannot be repeated, which actually limits the method to analytical applications.

More promising approaches for liquid-state NMR are (*i*) shuttle-DNP and (*ii*) high-field (HF) DNP, recently reviewed by Griesinger et al (2012). In the first method, the sample is polarized at lower field (e.g. 0.3 T or lower), and pneumatically shuttled into an NMR probe (the Göttingen design) or transferred with the entire DNP probe in to higher field (7 or 14 T) for NMR signal detection (the design in FU Berlin). The transfer times are as short as 40 ms or 0.3 s for the respective approaches. The coherence losses can be further minimized by increasing B_0 field between DNP polarizer and NMR field, e.g. by constructing two-center magnets or by employing neodymium transmission lines (as in Lausanne). Enhancement factors of 10 have been reported, however, the low-field Boltzmann polarization of electrons influences it adversely. The advantage of the shuttle-DNP is that μL sample volumes and aqueous solutions can be used due to smaller sample heating in DNP polarizer at low field.

In the alternative approach, HF-DNP, irradiation occurs at NMR field. Consequently, only nL volumes and non-polar solvents can be used to avoid high dielectric losses and sample degradation. Enhancements ϵ ∼20 were obtained at B_0=9.4 T, however, the applications are

limited to size-restricted samples, yielding a small filling factor of NMR probe. In conclusion, liquid-state DNP is close to approach biomolecular applications; two methods ensure high-resolution and repetition rates of seconds, both crucial for multidimensional NMR.

Noteworthy is also the recently proposed surface enhanced DNP NMR spectroscopy (Lesage et al 2010). Here, the silica nuclei present on the surface of a solid material are selectively hyperpolarized by biradical molecules. As they do not penetrate bulk material, the clear and strong NMR signal is emitted only from the surface. The method is particularly promising in studies of catalysts and porous materials (Griffin 2010).

1.4.2. Attenuated transverse relaxation

CSA/DD interference At high magnetic fields the contribution of CSA to relaxation of some nuclei (e.g. ^{15}N, C') becomes significant, and similar in magnitude as from dipolar interactions. In favourable cases, DD and CSA fields may be strongly correlated due to rigidity of respective structural motif. E.g. in a H^N-^{15}N moiety, the axially symmetric CSA tensors of both H^N and ^{15}N are aligned parallel to N–H bond, causing a strong interference of DD and CSA fields. In these conditions, the relaxation rates of individual transitions $S_{12}^{\pm}$ and $S_{34}^{\pm}$ may be significantly different:

$$R_{12} = (p-\delta_S)^2(4J(0)+3J(\omega_S)) + p^2(J(\omega_I-\omega_S)+3J(\omega_I)+6J(\omega_I+\omega_S)) + 3\delta_I^2 J(\omega_I) \quad (1.24)$$

$$R_{34} = (p+\delta_S)^2(4J(0)+3J(\omega_S)) + p^2(J(\omega_I-\omega_S)+3J(\omega_I)+6J(\omega_I+\omega_S)) + 3\delta_I^2 J(\omega_I) \quad (1.25)$$

where $J(\omega)$ is a spectral density function,

$$p = \frac{1}{2\sqrt{2}}\gamma_I\gamma_S\hbar r_{IS}^{-3}$$

is due to dipolar field of spins separated by r_{IS}, and

$$\delta_S = \frac{1}{3\sqrt{2}}\gamma_S B_0 \Delta\sigma_S, \quad \delta_I = \frac{1}{3\sqrt{2}}\gamma_I B_0 \Delta\sigma_I$$

is a contribution from CSA, where $\Delta\sigma_X$ stands for a difference between axial and perpendicular components of CSA tensor of spin X. As can be appreciated from eq. 1.24, whenever p is comparable to δ_S, the dominant term proportional to $J(0)$ may cancel for one of transitions. Pervushin et al (1997) showed that this principle can be utilized for N–H moieties, where $\Delta\sigma_H = -16\,ppm$ and $\Delta\sigma_N = -160\,ppm$. Optimal conditions for relaxation difference, i.e. $p \approx \omega_S$ and $p \approx \omega_I$, are different for H^N and ^{15}N, and occur at B_0 field corresponding to ^{1}H freq. of approx. 1.2 and 0.9 GHz, respectively.

To exploit *transverse-relaxation optimized spectroscopy* (TROSY) principle mixing between transitions, by either ^{15}N decoupling or H^N refocusing pulses during $t_2(^1\mathrm{H})$ or $t_1(^{15}\mathrm{N})$, should be avoided. Instead, only most slowly relaxing coherences are observed, and selection may be obtained by use of dedicated ST2-PT block (see par. 1.3.2). Inherent drawback of the approach is thus a nominal loss of 50% of signal, which may not be compensated by slower relaxation for small and medium sized macromolecules.

An interesting feature of TROSY is the possibility to utilize steady-state ^{15}N polarization. The actual contribution is larger than expected from the gyromagnetic ratios γ_N and γ_H since heteronuclei feature shorter T_1 at high fields (Pervushin et al 1998b). This effect is even more pronounced for aromatic ^{13}C-^{1}H-TROSY in nucleic acids, where steady-state ^{13}C polarization contributes up to 50% of initial magnetization (Brutscher et al 1998).

TROSY is most efficient if other sources of relaxation such as dipolar interactions with remote ^{1}H spins are minimized. Therefore, TROSY performs best for ^{2}H,^{15}N- or triple-labelled (H^N back-exchanged) proteins. However, it is not a prerequisite for TROSY for imino grops in nucleic acids due to considerably lower ^{1}H spin density (Farjon et al 2009).

CSA/DD interference has also proved beneficial for aromatic ^{13}C-^{1}H moieties in proteins (Pervushin et al 1998a) and nucleic acids (Brutscher et al 1998). In this case, however, the negligible CSA of aromatic ^{1}H spins prevents effective field cancellation, and TROSY selection is employed only in ^{13}C dimension. Also, evolution due to ^{13}C-^{13}C couplings should be refocused in a constant-time evolution, or by dilution of ^{13}C spins (Brutscher et al 1998). Otherwise, little gain in resolution can be obtained.

TROSY represents one of the major recent breakthroughs in biomolecular NMR. Nearly all experiments incorporating ^{15}N evolution period have been modified to utilize CSA/DD interference. They have been comprehensively reviewed by Zhu and Yao (2008).

DD/DD interference For isolated CH_3 groups, e.g. in otherwise perdeutrated proteins, significant cross-correlation between dipolar interactions within the moiety can also be exploited for relaxation attenuation. In particular, destructive interference of ^{1}H-^{1}H and ^{1}H-^{13}C dipolar fields cause the ∼9-fold difference in relaxation of two inner w. r. t. two outer lines of CH_3 multiplet. Since the effect derives from dipolar interactions, it is field-independent (unlike CSA/DD interference). Tugarinov et al (2003) showed that HSQC prevents from constructive use of the significant difference of relaxation of individual transitions, since slowly- and fastly-relaxing coherences are interconverted by ^{1}H π pulse. In contrast, ^{1}H-^{13}C ZQ/DQ coherence records three transitions, where the central line has nearly vanishing relaxation rate. The outer lines are interconverted by π ^{1}H pulse, and superimpose onto the narrow central line in HMQC. However, the linewidth of the latter remains affected. Therefore, HMQC can be employed „as is" (without any modification) for deuterated large systems since the troublesome lines are broadened beyond detection leaving only narrow components. The so-called *methyl*-TROSY effect revolutionized NMR of very large oligomeric proteins, yielding a tool for investigation of individual CH_3 sites in systems of molecular mass ranging to 1 MDa.

^{1}H decoupling Liu et al (2003) made an interesting observation that the relaxation of inner components of ^{13}C quartet in the methyl groups in fully protonated proteins is dominated by the proton flip-flop mechanism. Therefore, it is beneficial to transform the anti-phase $C_{xy}H_z$ into the in-phase C_{xy} operator, and prevent the evolution of couplings by ^{1}H decoupling. For example, the replacement of refocusing $\pi(^1\mathrm{H})$ pulse by decoupling during the CT evolution of ^{13}C lead to a sensitivity improvement in CH_3-detected HCCH-TOCSY (Yang et al 2004).

Similarly, H^N decoupling during storage of the ^{15}N in-phase transverse magnetization prevents from signal losses due to the exchange of H^N with solvent.

Multiple-quantum coherences It has been long recognized that in the slow-tumbling limit, the relaxation of ^{1}H-^{13}C ZQ and DQ coherences is not affected by dipolar interaction between involved spins (Griffey and Redfield 1987). In the case of H^α and C^α discussed by Grzesiek and Bax (1995), the respective relaxation rates of ^{1}H and ^{13}C SQ (average for in-

and anti-phase terms), and MQ coherences, are:

$$R_{\mathrm{SQ(H)}} = \frac{1}{5}J(0)\left(d_{HC} + \sum_S d_{HS} + \frac{5}{4}\sum_I d_{HI}\right) \tag{1.26}$$

$$R_{\mathrm{SQ(C)}} = \frac{1}{5}J(0)\left(d_{HC} + \sum_I d_{IC} + \frac{1}{4}\sum_I d_{HI}\right) \tag{1.27}$$

$$R_{\mathrm{MQ(C-H)}} = \frac{1}{5}J(0)\left(\sum_I d_{IC} + \frac{5}{4}\sum_I d_{HI} + \sum_S d_{HS}\right) \tag{1.28}$$

where CSA and cross-correlation terms were neglected, $d_{ij} = (\hbar\gamma_i\gamma_j r_{ij}^{-3})^2$, and summations extend over all protons, I, and carbons S, other than the proton, H, and carbon, C, involved in the coherence. As can be supposed from the absence of the term d_{HC} in 1.28, the MQ coherence should relax considerably slower than SQ coherences. Indeed, Grzesiek and Bax (1995) estimated that $R_C/R_{MQ} = 1.7$ and $R_H/R_{MQ} = 2.4$ for H^{α}-C^{α} moiety without H^N. The effect, referred to as multiple-quantum line narrowing, was extensively exploited in protein (Grzesiek et al 1995; Shang et al 1997; Swapna et al 1997; Gschwind et al 1998) and RNA experiments (Marino et al 1997; Fiala et al 1998; Hu and Jiang 1999). It should be emphasised that in these applications the deleterious effects of homonuclear couplings were either avoided by random fractional deuteration, or refocused in CT periods or by selective ^{1}H pulses.

Selective protonation and fractional random deuteration The deuteron has a 6.5-fold smaller gyromagnetic ratio than the proton: $\gamma_H/\gamma_D \approx 6.5$. Therefore its contribution to dipolar relaxation of other nuclei (either ^{1}H or heteronuclei) is lower by a factor of

$$\left(\frac{\gamma_H}{\gamma_D}\right)^2 \frac{I_H(I_H+1)}{I_D(I_D+1)} \approx 16 \tag{1.29}$$

where $I_H = \frac{1}{2}$ and $I_D = 1$ are spin quantum numbers of ^{1}H and ^{2}H (Cavanagh et al 2007, §9.1). Calculations and measurements showed that perdeuteration has a dramatic effect on T_2 of C^{α} (Yamazaki et al 1994) and H^N (Markus et al 1994), while a little difference is observed for ^{15}N and C', for which a dominant relaxation mechanism is CSA. The increase of T_2 with deuteration level d is steepest for H^N, however, above d=0.8 it is also accompanied by rapid increase of T_1. Therefore, the optimal condition for H^N-detected experiments is a high, but not complete deuteration (Nietlispach et al 1996). A suite of „out-and-back" experiments for sequential assignment based on H^N, ^{15}N, C^{α} and C' chemical shifts was proposed (Yamazaki et al 1994). Perdeuterated and selectively reprotonated proteins (see par. 1.3.1) obviously exhibit the largest lifetimes of spin coherences, however, recycle delays have to be optimized to account for longer T_1.

Shared-time and constant-time evolution Shared time evolution[3] partially utilizes the coherence transfer delay (usually $2\tau = 1/(2J)$) as an evolution period. This can be implemented as a maximum possible contraction $\xi = \tau\frac{t}{t_{\max}}$ of both delays flanking π pulse on the evolved nucleus, which does not affect evolution of chemical shift or any heteronuclear

[3]Commonly, the term „semi-constant" time is also used, which should rather be avoided since a time delay can only be *constant* or *variable*.

couplings. Effectively, the overall time with transverse magnetization decreases by 2ξ, which amounts to reduction of apparent relaxation rate of evolved coherence:

$$s(t) = s(t_0)\exp(-R_2^*(t-2\xi)) = s_0\exp\left(-R_2^*\left(t-2t\frac{\tau}{t_{max}}\right)\right) = s_0\exp\left(-\underbrace{R_2^*\left(1-\frac{2\tau}{t_{max}}\right)}_{R_2^{\text{eff}}}t\right)$$

As an effect, the signal loss during evolution period is reduced by the factor of:

$$\epsilon = \frac{(R_2^{\text{eff}})^{-1}(1-\exp(-R_2^{\text{eff}}t_{\max}))}{(R_2^*)^{-1}(1-\exp(-R_2^*t_{\max}))}$$

The apparent attenuation of relaxation is maximal for $t_{\max} \leq 2\tau$, which corresponds to *constant-time* evolution with $R_2^{\text{eff}} = 0$ and enhancement factor of

$$\epsilon = \frac{t_{\max}}{(R_2^*)^{-1}(1-\exp(-R_2^*t_{\max}))}$$

Recently, the author proposed to utilize both coherence transfer delays already present in ^{1}H-^{13}C HMQC for shared-time ^{1}H evolution. The simple modification of the block, shown in Fig. 1.9, was estimated to improve sensitivity by up to 18%, which was also verified experimentally on 4D ^{13}C,^{15}N-edited NOESY of MBP (Stanek et al 2012). Similar idea was exploited for this experiment by Xu et al (2006), however, with additional π pulses. The optimized HMQC (with necessary adjustments) can be also incorporated into 4D ^{13}C,^{13}C-edited NOESY for RNAs (Stanek et al 2013b) and proteins.

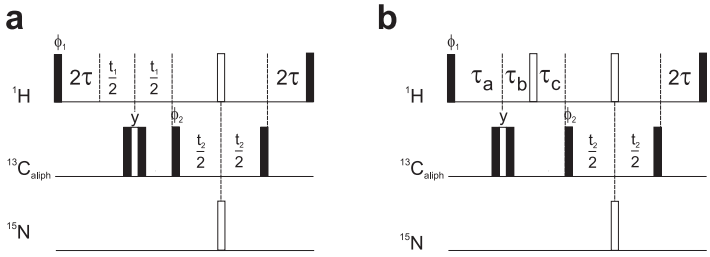


Fig. 1.9 The comparison of the conventional HMQC (a), and the recent optimization (b). For shared-time evolution the delays τ_a, τ_b and τ_c are $t/2$, $t/2\cdot(1-4\tau/t_{\max})$ and $2\tau(1-t/t_{\max})$, respectively. If $t_{\max} < 4\tau$, the delays are $\tau_a = t/2$, $\tau_b = 0$ and $\tau_c = 2\tau - t/2$.

1.4.3. Hardware improvements

Static field strength B_0 Increasing field strength has linear effect on Boltzmann equilibrium polarization, while higher frequency of FID signal scales current induced in the coil, but also enhances electronic noise as $\sqrt{\omega}$ (Keeler 2005). In effect, the nominal increase of S/N $\propto B_0^{3/2}$. Additionally, spectral widths scale linearly with B_0, which, in absence of sampling restrictions, enables to enhance spectral resolution. Also strong coupling effect, e.g. for ^{1}H in aromatic rings, are diminished. These advantages are clearly the driving force for constantly higher B_0 field of modern NMR systems, with the strongest one of 23.5 T installed in 2009 in ENS Lyon. In addition, it is predicted that the optimal line-narrowing effect of CSA/DD interference for H^N should occur at $B_0 \approx 1.1$ GHz (Pervushin et al 1998b), and such superconducting magnets are expected to be soon commercially available.

However, increasing bandwidths cause notorious troubles with uniform excitation, degrade performance of isotropic TOCSY mixing sequences, and pose high demands on heteronuclear decoupling. Also, as explained in sec. 2.1.1, due to sampling requirements, spectral resolution in indirect dimensions is not effectively improved. Further complications arise due to

attenuated T_1 relaxation, which implies longer recycling delays, and enhanced T_2 relaxation of nuclei which characterize a dominant contribution of CSA mechanism, e.g. ^{13}CO, aromatic ^{13}C or ^{15}N. One should also consider increasing technical difficulties with frequency synthesis and generation, and transmission of electromagnetic waves for $\omega > 1\,\mathrm{GHz}$, which may affect the performance of high-field systems.

Cryogenically cooled probes Sensitivity, being the critical parameter of NMR spectrometers, is predominantly affected by incoming signal strength I_S, and electronic (thermal) noise generated in the rf. coil and preamplifier, I_N. In general, the S/N ratio is governed by multiple factors (Cavanagh et al 2007, §3.1):

$$S/N \propto \frac{I_S}{I_N} \propto \frac{N\gamma_{\mathrm{ex}}\gamma_{\mathrm{det}}^{3/2}B_0^{3/2}}{\sqrt{\Delta\Omega(R_c(T_c+T_{\mathrm{amp}})+R_s(T_{\mathrm{amp}}+T_s))}} \tag{1.30}$$

where, in addition to obvious: number of spins N, the gyromagnetic ratios of excited and detected nuclei ($\gamma_{\mathrm{ex}}, \gamma_{\mathrm{det}}$), and the static field B_0, important role is played by the filter-windowed bandwidth open to preamplifier $\Delta\Omega$, the coil resistance R_c and temperature T_c, the effective temperature of noise in preamplifier T_{amp}, and the sample temperature T_s and resistance induced in the coil by the sample, R_s. It can be appreciated from the equation that cooling the coil and preamplifier (which has to be integrated into the probe) can result in dramatic improvements of sensitivity. The cryogenically cooled probes using gaseous helium were introduced by Bruker for commercial use in 1999, and now operate at ∼20 K. For low-conductivity buffers the S/N gains as large as a factor of four are feasible (Kovacs et al 2005). Recently, spectrometer vendors introduce less expensive nitrogen-cooled probes, which enable smaller, but still appreciable enhancements of factor of 2-3.

However, the side effect of lowering T_c and T_{amp} is the increased dependence on sample resistance R_s which cannot be eliminated due to the fixed sample temperature T_s. Effectively, the obtained gains vary greatly on the conductivity of buffer used. Since a very salty buffer can offset most benefits of a cryo-probe, alternatives such as use of small-diameter or non-cylindrical probes have been proposed (Cavanagh et al 2007). Noteworthy, cryo-probes usually feature lower filling factors because of necessary thermal isolation between the coil and sample, and possess limited capabilities of only z-axis PFGs. Another notorious problem occurs due to amplified radiation damping, and consequently, degraded performance of solvent signal suppression (Cavanagh et al 2007, §3.7).

The appearance of double (^{1}H,^{13}C)-receiver consoles, introduced by Agilent/Varian in 2006 is also noteworthy (see sec. 2.3.3).

Ultra-fast MAS In solid-state NMR (ssNMR) the major source of signal dephasing, or homogeneous line broadening, are dipolar, CSA and quadrupolar fields. Rapid spinning at the so-called magical-angle (MAS), ∼54.7° w. r. t. B_0 field, is capable of averaging those interactions[4] to a certain extent, similarly as effects from molecule tumbling in solution. The dipolar mechanisms scale with gyromagnetic ratios of involved nuclei, and its less demanding to average them by MAS for ^{13}C and ^{15}N spins, which until recently dominated in ssNMR.

With the advent small-diameter rotors (0.7 mm) and MAS probes achieving spinning rate of 60 kHz[5] the ^{1}H-^{1}H dipolar interactions can be decoupled sufficiently effective to record high-resolution ^{1}H-detected spectra (Zhou et al 2007a; Bertini et al 2010; Laage et al 2009), see Fig. 1.10. Obviously, this provides a substantial sensitivity improvement due to larger

[4] all having traceless tensor representations

[5] The first commercial 100 kHz MAS probe to be installed in CNRS Lyon late fall of 2014

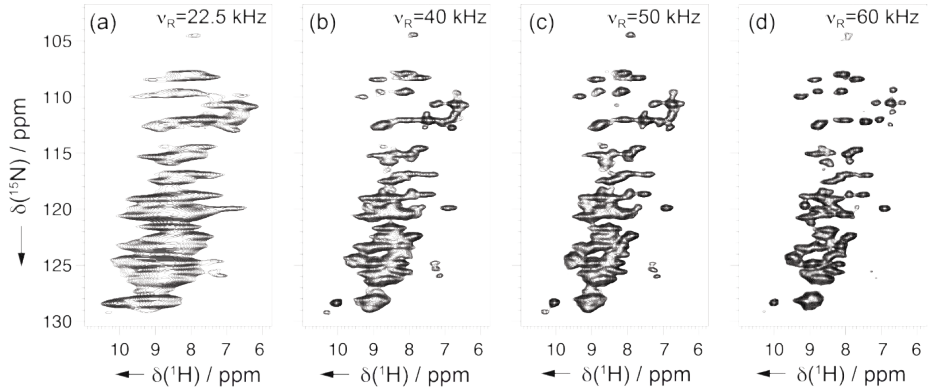


Fig. 1.10 The improvement in resolution of H^N-detected ^{15}N-^1H correlation spectra of *E. Coli* [^{13}C,^{15}N]-SSB (4×18 kDa) recorded at 800 MHz spectrometer and MAS rate of 22.5 (*a*), 40 (*b*), 50 (*c*), and 60 kHz (*d*). Reprinted from Marchetti et al (2012).

gyromagnetic ratio of ^1H. Moreover, the increase of B_0 field increases the ^1H coherence lifetimes more than linearly owing to truncation of dipolar terms in the Hamiltonian (Marchetti et al 2012). Contrary to the heteronuclear case, ^1H-^1H interactions are subject to complete averaging at yet inaccessible MAS rates (> 200 kHz), therefore, ^1H-detected NMR must be aided by deuteration. While at moderate MAS rates amide sites must be only partially reprotonated to dilute proton bath and preserve high-resolution ^1H spectra (Chevelkov et al 2006; Hologne et al 2006; Akbey et al 2010), 60 kHz MAS at B_0=25 T allows for 100% protonation of exchangeable amide protons and proportional sensitivity improvement (Knight et al 2011).

Additionally, longer coherence lifetimes enable *J*-coupling based coherence transfer also in ssNMR (Lesage et al 1999; De Paëpe et al 2003). A suite of H^N-detected experiments for sequential assignment of resonances in proteins, and applications to perdeuterated and fully protonated non-crystalline samples were recently reported (Knight et al 2012; Marchetti et al 2012; Zhou et al 2007b). These protocols resemble experiments performed for over 2 decades in liquid-state NMR, and, in liason with recent sensitivity and resolution enhancement of ssNMR, open avenues to study structure and dynamics of large proteins. Of the particular relevance is that (*i*) the linewidths observed in ssNMR are independent of molecular mass, and (*ii*) non-crystalline natively highly-hydrated samples can be used, with the use of ultracentrifugation (Bertini et al 2011; Gardiennet et al 2012).

Chapter 2

Accelerated acquisition of NMR spectra

2.1. Breakthrough in indirect sampling

2.1.1. Coupling between the line width and spectral window

To understand the interplay between sampling rate and line widths which results from uniform sampling, we have to resort to classical Fourier analysis. Let s_k be K samples of continuous signal $s(t)$ at (arbitrary) sampling points t_k:

$$s_k = s(t_k),\ k = 1, 2, \ldots, K \tag{2.1}$$

One can formally express the sampling process as multiplication by a step function $\Pi(t)$

$$\Pi(t) = \begin{cases} 1 & |t| < t_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

which mimics signal truncation, and infinite train of Dirac delta functions (Dirac comb):

$$\text{Ш}(t) = \sum_{k=-\infty}^{+\infty} \delta(t - t_k) \tag{2.3}$$

Discretely sampled signal is thus

$$d(t) = \sum_{k=1}^{K} s_k \delta(t - t_k) = s(t)\text{Ш}(t)\Pi(t) \tag{2.4}$$

where we assume the casuality of the signal, i.e. $s(t) = 0$ for $t < 0$. According to Convolution Theorem (Bracewell 2000, §6), its FT is a convolution of underlying Fourier representations:

$$D(\omega) = \mathcal{F}[d(t)] = \mathcal{F}[s(t)] * \mathcal{F}[\text{Ш}(t)] * \mathcal{F}[\Pi(t)] \tag{2.5}$$

Fig. 2.1*a-c* illustrates these components and subsequent convolutions. FT of step function can be simply derived

$$\Pi(\omega) = \int_{-t_{\max}}^{+t_{\max}} \exp(-i\omega t)\mathrm{d}t = 2\,t_{\max}\frac{\sin(\omega t_{\max})}{\omega t_{\max}} = 2\,t_{\max}\,\mathrm{sinc}(\omega t_{\max}) \tag{2.6}$$

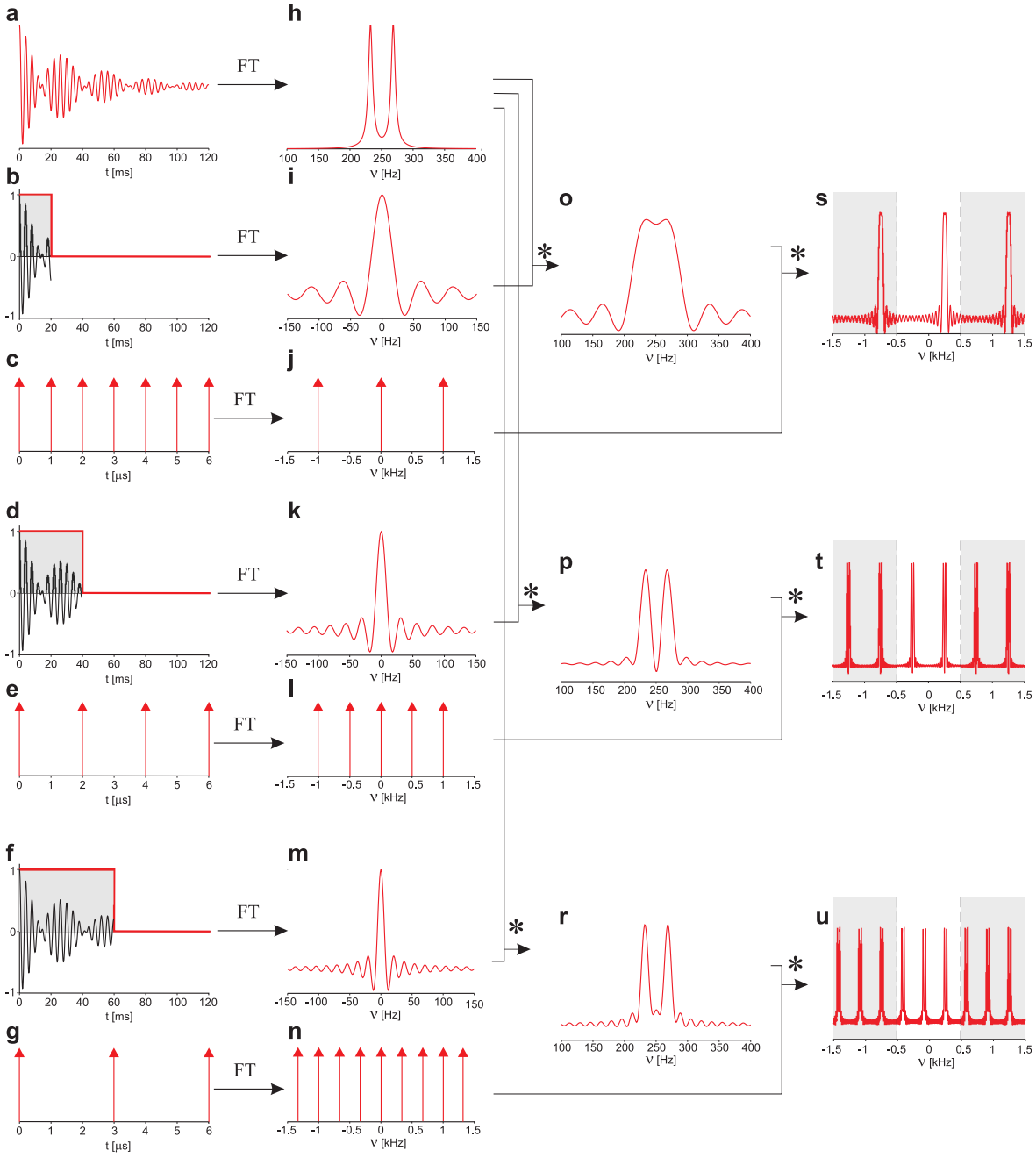


Fig. 2.1 Coupling between spectral resolution and folding (aliasing): *a* A synthetic signal comprising two components of frequency differing by 35 Hz and T_2^*=40 ms. b, d, f Step functions that mimic signal truncation at 20, 40 and 60 ms, respectively. *c, e, g* Trains of Dirac delta functions which formally represent discrete sampling. *h-n* Continuous Fourier Transformations of signals *a-g*, respectively. Extended sampling results in a narrowing of sinc profiles (*i, l, m*), however, this is accompanied by a lower sampling rate and, effectively, an increased aliasing in Fourier domain (*j, l, n*). Convolutions of spectra (denoted with a star) reveal the effect of signal truncation on a separation of doublet lines (*o, p, r*). Discrete Fourier Transformations of truncated signals (*s-u*) illustrate the aliasing phenomenon: additional copies of a resonance appear in the central 1 kHz spectral window.

Sinc-shaped profile of $\Pi(\omega)$ causes line broadening. Its full-width-at-half-height, and distance between maximum and the first null point are $\Delta_{\mathrm{FWHH}} \approx 1.26\pi\,(t_{\max})^{-1}$, and $\Delta_0 = \pi(t_{\max})^{-1}$,

respectively. Clearly, the more severe truncation of the signal, the broader are the spectral lines (cf. Fig. 2.1). However, the relative height of the first lobes is invariant of $t_{\max}$:

$$\begin{array}{l} \Pi(\omega = \frac{3}{2}\frac{\pi}{t_{\max}})/\Pi(\omega = 0) = -\frac{2}{3\pi} \approx -0.212 \\ \Pi(\omega = \frac{5}{2}\frac{\pi}{t_{\max}})/\Pi(\omega = 0) = \frac{2}{5\pi} \approx 0.127 \end{array} \tag{2.7}$$

Noteworthy is the magnitude of truncation artefacts, which poses serious difficulties for interpretation of some spectra, e.g. close to diagonal regions of NOESY. The commonly applied *apodization* (damping) functions decrease *sinc* wiggles, however, at the cost of further line broadening.

The sampling function, $\text{Ш}(t)$, has closed analytical form of FT only if samples are equidistant, $t_k = k\,\Delta_t$, i.e. for linear (uniform) sampling:

$$\text{Ш}(\omega) = \sum_{n=-\infty}^{+\infty} \delta(\omega - \omega_n), \quad \text{where } \omega_n = \frac{2n\pi}{\Delta_t} \tag{2.8}$$

As a consequence of eq. 2.5, DFT spectrum of uniformly sampled signal comprises an infinite number of copies of the continuous spectrum, separated by $\Delta_\omega = 2\pi(\Delta_t)^{-1}$. This phenomenon is commonly referred to as *aliasing* or *folding* (cf. Fig. 2.1*s-u*). One can also explain it by noting, that a signal of frequency $\omega + \Delta_\omega$ yields identical values at every measured point as oscillatory function of ω:

$$\exp(i(\omega + \Delta_\omega)t_k) = \exp(i\omega t_k + i\Delta_\omega k\Delta_t) = \exp(i\omega t_k + 2k\pi) = \exp(i\omega t_k) \tag{2.9}$$

Aliasing determines the spectral window that is correctly reproduced:

$$\text{sw} = \Delta_\omega = \frac{2\pi}{\Delta_t} \tag{2.10}$$

Combination of eq. 2.10 and line width, $2\omega_0$, leads to

$$\text{sw} = \frac{2\pi}{\Delta_t} = \frac{2\pi}{\Delta_t}\frac{K}{K} = 2K\frac{\pi}{t_{\max}} = K\omega_0 \tag{2.11}$$

which *explicitly* expresses the coupling between line widths and assumed spectral window. Thus, resolution can only by improved by increasing number of equidistant samples, K.

2.1.2. Limited sampling: truncation and sparse sampling

In the biomolecular NMR indirect sampling is an inherently slow process. Since biomolecules exhibit unfavourable relationship between transverse and longitudinal relaxation rates (the latter additionally increasing with B_0), vast majority of experimental time (80-90%) is spent on recovery of steady-state magnetization (Fig. 2.2). To avoid saturation, the optimal recycle delay should be $\approx 1.25T_1$, thus $d_1 \approx$ 1s. Also, if high-power decoupling is employed, a minimum safe duty cycle must be maintained to prevent from damage of probe circuits. Additionally, a minimal 4- or 8-step phase cycling, as well as 2^{D-1} quadrature components are performed. As a result, typical sampling rate in biomolecular NMR is of order of 1-2 points per minute.

A very general rule for indirect sampling is to match max. evolution time to signal envelope. A reasonable trade-off between resolution and sensitivity is $t_{\max} \approx 1.26T_2^*$, since

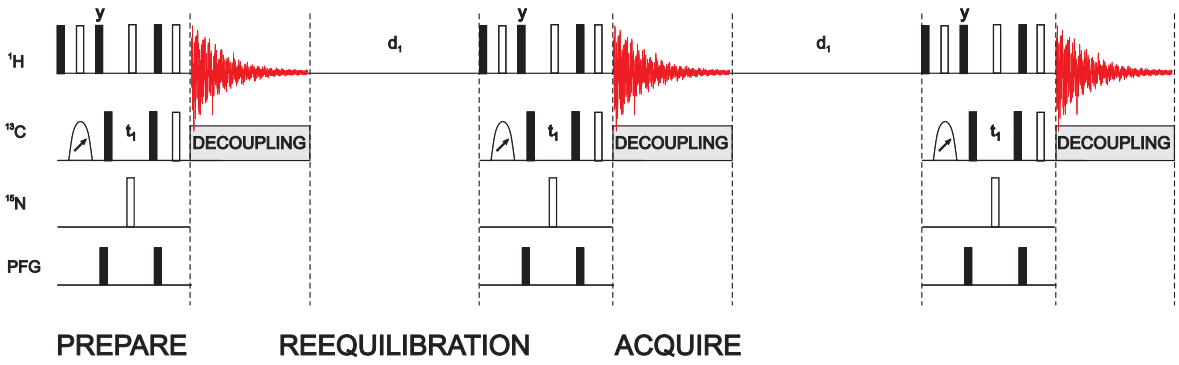


Fig. 2.2 A general scheme of indirect sampling: acquisition periods (up to 100 ms) must be preceded by long T_1-recovery delays ($d_1 \gtrsim 1$ s), and signal preparation periods (pulse sequence, 10-200 ms).

time-domain sensitivity gently decreases beyond this limit (Rovnyak et al 2004b). If *aliasing* is to be avoided, the required number of samples for an D-dimensional experiment is thus

$$N = \prod_{i=1}^{D-1} N_i = \prod_{i=1}^{D-1} \mathrm{sw}_i \cdot t_{\max i} = \prod_{i=1}^{D-1} \Delta\Omega_i \, B_0 \cdot 1.26 \, T^*_{2,i} \tag{2.12}$$

where $\Delta\Omega_i$ is chemical shift dispersion of the coherence evolved in ith dimension. If we assume for simplification that relaxation rates are field-independent, we immediately conclude that

$$N \propto (B_0)^{D-1} \tag{2.13}$$

i.e. the required number of samples increases *polynomially* (linearly, quadratically, etc.) with static field strength, and *exponentially* with dimensionality. As emphasised by Rovnyak et al (2004b), these requirements can easily be met only in two-dimensional spectroscopy, and nearly all routinely used 3D protein experiments would require impractical experimental times.

Obviously, experimental time is not only limited by availability and maintenance cost of NMR hardware, but also by limited stability of biological samples which may undergo aggregation, degradation or other reactions. In principle, there are three solutions to maintain reasonable duration of an experiment. One can deliberately increase a dwell time Δ_t, which leads to higher $t_{\max}$, and *controllable* aliasing (Jeannerat 2003, 2007). In this approach, the spectral window as well as line widths are decreased simultaneously, and spectral crowding remains unaffected. In contrast, spectral resolution is truly improved. However, multiple folding unavoidably leads to signal overlap, hence this solution is not a general remedy.

The second, and most commonly applied choice is signal truncation. As argued by Rovnyak et al (2004b), in typical 3D experiments performed at high-field spectrometers $t_{\max} \approx 0.5 T_2^*$ is achieved at best, thus resolving potential of these equipments is not exploited. In fact, to keep constant experimental time, both line widths and *spectral widths* are increased linearly as B_0, and spectral resolution (in ppm) remains unchanged. Of a great concern are also prominent truncation artefacts which appear as *sinc* sidelobes. These can be particualrly detrimental as demonstrated in Fig. 2.3. Noteworthy, truncation artefacts are well-localized (proximal to peaks). Increased number of samples causes their compression towards peak centre, without change in intensity. Signal extrapolation methods such as linear prediction (Koehl 1999), maximum entropy reconstruction (Davies et al 1988) or lineshape deconvolution using CLEAN (Keeler 1984) were proposed to tackle truncation artefacts. Nowadays, these solutions are commonly applied, however, considerable criticism was

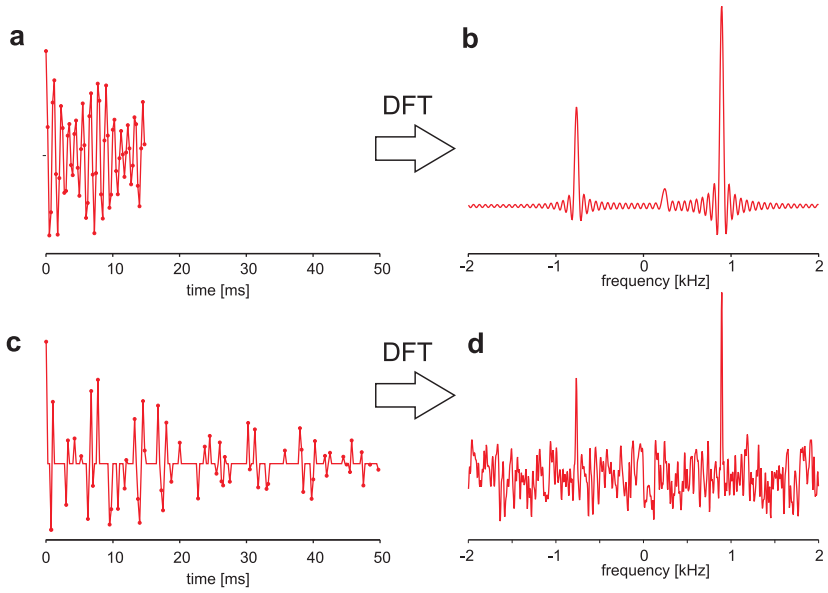


Fig. 2.3 Two approaches for sparse sampling: (*a*) signal truncation at $t_{\max} \leq T_2^*$ and (*b*) non-uniform (random) sampling to achieve $t_{\max} \gtrsim T_2^*$. Both solutions generate sampling artefacts: (*b*) *sinc* wiggles, most pronounced in the proximity of peaks, and (*d*) non-local pseudo-random noise. Simulated signal contains 3 components of $T_2^* = 33$ ms.

raised regarding true capabilities of those methods (Stern et al 2002). Merely, these methods suppress truncation artefacts by signal extrapolation, however, observed line narrowing does not correspond to real resolution enhancement. In other words, there are no reliable methods for signal extrapolation from severely truncated signals.

A very general alternative is sparse random sampling (Fig. 2.3c), which enables true resolution enhancement. However, non-local spectral artefacts appear in spectra, which resemble pseudo-random noise. Noteworthy, point response is deterministic and it is in principle possible to deconvolve it from spectral lines. Such methods are introduced in Section 2.4. Needless to say that random sampling artefacts exhibit quite distinct behaviour compared to truncation artefacts: increased number of samples decreases their intensity without affecting their (flat) spectral distribution. More features of NUS artefacts are discussed in Sec. 2.4.2.

2.2. Non-uniform sampling strategies

A number of various sampling schemes were proposed for multidimensional NMR. Recent review of Kazimierczuk et al (2012) discusses different approaches and associated artefact patterns. A detailed coverage of this topic is certainly beyond the scope of this thesis. Hence, only selected issues will be discussed here.

2.2.1. Deterministic and random sampling schemes

Non-linear sampling dates back to Barna et al (1987), who proposed to select n samples from N Cartesian nodes according to exponentially decaying density,

$$D(t) = \exp(-kt) \tag{2.14}$$

where k is determined from normalization condition: $\int_0^N D(t)\mathrm{d}t = n - 1$. Based on the rounded values of the integral,

$$I_j = \left\lfloor \int_0^j D(t)\mathrm{d}t \right\rfloor \qquad \text{for } j = 1, \ldots, N \tag{2.15}$$

one selects the first sampling point j which satisfies $I_j = l$ for $l = 0, \ldots, n-1$. As can be appreciated from the description, the resulting schedule is entirely deterministic, and depends only on n and N. Noteworthy is the first application of two-dimensional non-linear sampling to 3D spectroscopy in conjunction with MaxEnt processing (Robin et al 1991). The method of Barna et al (1987) was employed in both indirect ^{1}H dimensions, however, the sampling process was only simulated by discarding a part of already measured data. The so-called *exponential sampling* was later adapted to reflect arbitrary decay rate (k), and also anti-phase signals in 2D COSY-type spectra (Schmieder et al 1993). Soon, Schmieder et al (1994) concluded that random selection of samples yields more uniform and lower intensity of NUS artefacts. In the presented example, only ^{15}N dimension of 3D ct-HNCO was sampled non-uniformly; this restriction was mainly practical as it allowed fast 1D MaxEnt processing. With the increase of computing power and development of efficient MaxEnt processing (Li et al 1998), these limitations lost their relevance, and random sampling in two-dimensional indirect domain was introduced (Rovnyak et al 2004a; Frueh et al 2006b). Also, multi-way decomposition was proved capable of processing such data (Orekhov et al 2001, 2003).

At that time deterministic approaches to sparse sampling, represented primarily by radial sampling, gained considerable popularity. This should be attributed to either (*i*) conceptually simple analysis, (*ii*) higher sensitivity and (*iii*) low computational costs of spectral reconstruction. It should be noted that projection spectroscopy was derived from well-known concepts of Accordion Spectroscopy, Reduced-Dimensionality, and back-projection employed in MRI, and was therefore easily accepted by NMR community. However, strong artefacts of back-projection reconstruction or ambiguities that may arise during APSY protocol were soon recognized. Kazimierczuk et al (2006a) showed that spiral sampling leads to significantly lower artefacts, and employed FT as a convenient tool to process arbitrarily sampled data. This opened up avenues to compare different sampling approaches. Soon, it was concluded that randomly sampled schemes are indeed superior to any deterministic ones due to absence of coherent artefacts (Kazimierczuk et al 2006b). Coggins and Zhou (2007) introduced sampling along concentric rings, and also concluded that randomization diminishes artefacts. The approach was proven beneficial with respect to radial sampling, and later adjusted to equispaced fine grid to enhance processing in FFT-CLEAN (Coggins and Zhou 2008). Also Pannetier et al (2007) proved that randomized schemes generate spectra which, in contrast to radially sampled ones, do not suffer from systematic artefacts. The issue was also investigated by (Mobli et al 2006), who compared radial, on-grid approximation to radial, and random two-dimensional sampling. Irrespectively of employed reconstruction method, either additive back-projection or MaxEnt, the ambiguities stemming from radial sampling hampered correct peak identification, suggesting that they are intrinsic to sampling scheme. Moreover, frequency errors or loss of resolution accompany intense artefacts of back-projection reconstruction. Convincing arguments in favour of random sampling were also given by Hoch et al (2008), who demonstrated effects of gradually increased *blurring* of radial sampling schemes, and that randomization is more efficient in decreasing NUS artefacts than adding projection. Methods based on projections such as APSY perform well for well-dispersed spectra, for which the use of high-dimensional spectroscopy seem nothing but overcomplication. For degenerate spectra of complex biomolecules many additional projections are required (Coggins et al 2005), and thus their time efficiency is questionable (Mobli et al 2006).

In conclusion, the extensive literature suggests that deterministic schemes such as radial or spiral sampling lead to localized, albeit strong artefacts. On the contrary, random sampling yields non-local artefacts of considerably lower intensity, which was shown superior in more demanding applications. Therefore, all experimental examples throughout this thesis utilize random sampling.

2.2.2. On- and off-grid random sampling

The introduction of generalized FT enabled processing of arbitrarily, also off-grid sampled data (Kazimierczuk et al 2006a,b). Apart from effective invalidation of the notion of *bandwidth* (Tarczynski and Allay 2004), it also provided flexibility for different optimizations of sampling schedules (Kazimierczuk et al 2007, 2008a) and simplified analytical derivation of properties of NUS artefacts (Kazimierczuk et al 2009). However, the actual relevance if off-grid sampling seems overestimated and this issue should be clarified.

Firstly, on-grid sampling facilitates the use of FFT (Cooley and Tukey 1965) by simple sorting and zero-filling of data which was not measured. For a data record of N complex points, it requires $(2N+1)\log_2 N$ floating-point multiplications, which is a large progress in comparison to quadratic complexity (N^2) of computation of Fourier sum (DFT). However, for very sparse data ($K \ll N$) FFT is actually slower than Fourier sum, which requires only $N \times K$ operations. Additionally, symmetry properties of DFT can be utilized to accelerate computation of real-valued spectrum by a factor of two. The real benefits of on-grid arrangement of data points are revealed for multidimensional transforms, and we adopt a mixed strategy of using the faster of FFT or DFT as explained in detail in Sec. 3.2.3.

Secondly, it was proven that restriction on bandwidth (i.e. a finite sw) leads to increase of S/A ratio (Stanek and Koźmiński 2010). The effect was verified using PSF for a signal sampled with constant number of points and bandwith, and a variable $t_{\max}$. Results shown in Fig. 2.4 suggest that the increase of S/A is only slight at low sampling density ($\theta < 0.3$). Noteworthy, S/A dependence for both on and off-grid sampling seems independent of dimensionality. However, for higher dimensionality and constant N, sw, and $t_{\max}$ in all dimensions, $\theta \to 0$. Effectively, random on- and off-grid sampling lead to a very similar S/A.

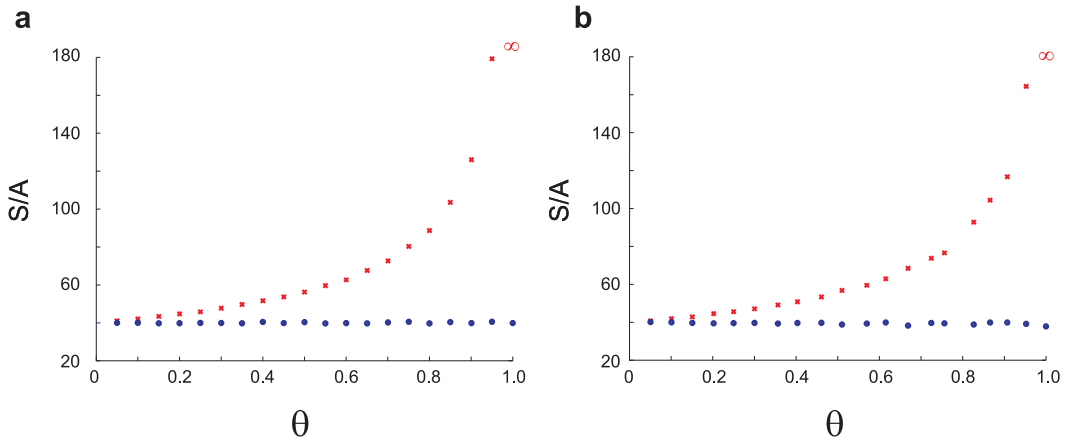


Fig. 2.4 Signal-to-artefact ratio measured in simulation of PSFs with constant number of sampling points, $N_{1D} = 800$ and $N_{2D} = 400$ for one- (a) and two-dimensional case (b), respectively. Spectral width(s) of 10 kHz was used, and relative sampling density was varied by changing the max. evol. time. 10 random sampling schemes were used for each measurement, and results averaged. For random off-grid sampling the S/A is constant, and agrees well with predictions based on eq. 2.48 (for 2D case, 2× quadrature components should be taken into account). In constast, sw restriction for on-grid sampling leads to increased S/A. Note that S/A$\to \infty$ as $\theta \to 1$ for the on-grid sampling.

Obviously, a prerequisite for on-grid sampling is the knowledge of expected bandwidths. However, in biomolecular NMR signals resonate in well-known regions and this does not imply any practical difficulties. It was also noted by Mobli et al (2006) that on-grid approximation of off-grid schedules does not lead to additional coherent artefacts. The non-trivial issue

of signal aliasing in the case of NUS was investigated by Maciejewski et al (2009). They emphasised that completely random selection (i.e., *off-grid*) of sample times does not afford infinite bandwidth, since effective sw is determined by the reciprocal of the greatest common divisor (GDC) of evolution times, which are always of a finite precision. Also, increased intra-band signal aliases (coherent NUS artefacts) can appear for some unfortunate choices of *on-grid* schedules, e.g. close to regularly spaced, but undersampling ones. This may be pronounced for very small subsets of a grid. To prevent this, some degree of oversampling was proposed. Effectively, random samples are likely to be relatively prime, and *intra-band* aliases are eliminated leading to decreased level of „sampling noise".

2.2.3. Optimized random sampling

The above mentioned findings of Maciejewski et al (2009) emphasise the relevance of *optimized* NUS schedules. These should (*i*) avoid regularity and (*ii*) large gaps between points, in particular (*iii*) at the beginning and end of the data (Hyberts et al 2010). There is however, no generally accepted choice of *optimal* sampling. Kazimierczuk et al (2007) suggested the use of *jittered* sampling to prevent from excessive inter-point distances, but later inclined to enforce more regular spacing using Poisson disks (Kazimierczuk et al 2008a). While the latter choice generates a larger region of decreased artefacts in the vicinity of peaks, it also introduces non-uniform distribution of noise at higher frequencies (*noise waves*), which could be misleading for automatic peak-picking algorithms. In addition, elevated noise is present parallel to frequency axes, which raises the question whether Poisson disks lead to over-regularized sampling schemes.

Hyberts et al (2010) assumed that imposing restrictions on random sampling should yield reproducible spectral reconstructions, i.e. of a quality independent on a specific choice of samples. It was proposed that gaps between points, k, should follow Poisson distribution of λ mean: $f(k;\lambda) = (\lambda^k \mathrm{e}^{-k})/(k!)$. Furthermore, to avoid large gaps at the beginning and end of data, sinusoidal variation of mean gap size was introduced (SPS): $\lambda = \Lambda \sin\theta$, where $\theta \in [0, \pi]$, $\theta \propto t/t_{\max}$, and Λ is an adjustable parameter to achieve required mean sampling density. Apodization may be imposed by simply restricting θ to $[0, \pi/2]$ interval. Hyberts et al (2010) showed that deviation (l_2-norm) of reconstructed spectrum (using Forward MaxEnt) from the reference was smallest for SPS, and contrary to unrestricted sampling, does not apparently vary upon a random generator seed. Effectively, there is a little risk of unfortunate choice of sampling schedule and final results are more reproducible.

2.2.4. Matched sampling

In fact, NUS was originally proposed to enhance sensitivity of 2D spectroscopy by matching sampling probability to a signal envelope, which is usually an exponential decay for in-phase signals (Barna et al 1987). Similar idea of collecting variable number of scans (*matched accumulation*) was also proposed for sensitivity enhanced 2D NOESY spectroscopy (Kumar et al 1991). In order to quantify possible sensitivity gains, we first derive S/N dependence for uniform sampling. In this case,

$$S/N \propto \frac{1 - \exp(-t_{\max}/T_2^*)}{\sqrt{t_{\max}/T_2^*}} \tag{2.16}$$

and the corresponding curve is plotted in Fig. 2.5*a*. The optimal sensitivity is achieved for $t_{\max}$ fulfilling the eq. $t_{\max} = T_2^* \log(2t_{\max}/T_2^* + 1)$, which is approximately $t_{\max} \approx 1.26\, T_2^*$. It can be inferred from the S/N dependence that while sampling up to $\approx 0.7\, T_2^*$ significantly

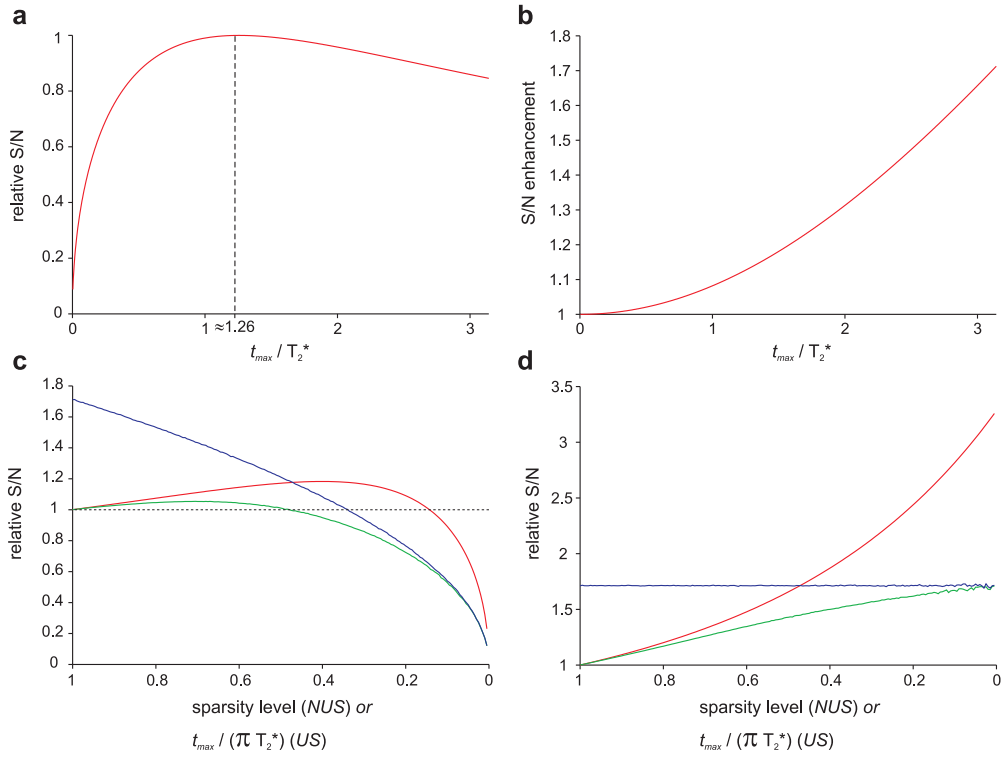


Fig. 2.5 (*a*) Signal-to-noise (S/N) ratio for uniform truncated sampling as a function of obtained max. evol. time. (*b*) S/N gain of matched exponential sampling compared to time-equivalent uniform sampling to the same $t_{\max}$. (*c*) Relative S/N of uniform (*red*), and matched on-grid exponential sampling with at most one transient per point (*green*) or without this restriction (*blue curve*). The simulation shows the sensitivity for sampling with various number of samples. For US the max. evol. time (truncation) is varied while in the case of NUS the sampling always extends up to $\pi\, T_2^*$, and its density is adjusted. (*d*) Sensitivity for time-equivalent sampling of an exponentially decaying signal. Color-coding as for panel *c* is used. The dependencies were derived from those shown in panel *c* by multiplication by $\sqrt{1/\theta}$, where θ denotes sparsity level, to account for accumulation of additional transients.

improves sensitivity, any point beyond optimal $t_{\max}$ degrades it. The aim of matched NUS is thus to transfer a fraction of samples from the late to early part of the signal. Rovnyak et al (2011) derived expression for S/N gain of matched exponential sampling with respect to uniform sampling, both truncated at $t_{\max} = \epsilon\, T_2^*$:

$$\eta_{\text{matched}} = \frac{\epsilon(1-\exp(-2\epsilon))}{2(1-\exp(-\epsilon))^2} \tag{2.17}$$

As can be appreciated for Fig. 2.5*b* very slight gains of approx. 8% are attainable for severely truncated data records ($t_{\max} = T_2^*$). However, if high-resolution spectra are of interest, i.e. $t_{\max} = 2\, T_2^*$ or $3\, T_2^*$, considerable sensitivity enhancements of 31 and 66% are obtained. Noteworthy, maximally resolved spectra are sometimes employed in biomolecular NMR as an alternative to increased dimensionality, which may entail prohibitive signal losses arising during additional coherence transfer periods. More commonly, a trade-off between sensitivity and resolution is made, and compromised sampling up to $\approx 1.5\, T_2^*$ is employed. Despite sensitivity gain is rather moderate in a single NUS dimension, the cumulative effect in high-dimensional experiment may be significant (Hyberts et al 2013). Additionally, time-savings are more impressive since they scale as η^2.

The discrete nature of on-grid sampling causes that perfect matched sampling is not feasible, and long ducts of linearly sampled data appear for dense NUS schedules ($\theta \gg 0$). As shown in Fig. 2.5*c* NUS converges to linear sampling for $\theta \to 1$, and actual S/N curve for NUS (*green curve*) is an intermediate between limiting cases of truncated uniform (*red*) and perfectly exponential distribution of samples (*blue*). Increasing sparsity in NUS schedules initially enhances sensitivity as points at high times are discarded, however, further reduction (beyond $\theta \approx 0.5$) degrades S/N since points at the beginning of the signal are less populated. If time-equivalent sampling is considered, i. e. sparsity is compensated by additional signal accumulation, NUS is always beneficial compared to *uniform* sampling as shown in Fig. 2.5*d*, and maximum gain of ≈71% is obtained for very sparse sampling ($\theta \to 0$).

Matched NUS sampling is commonly used in biomolecular NMR as a general strategy to enhance resolution without significant losses in sensitivity. As estimated by Rovnyak et al (2004b) only moderate 20% of peak intensity is typically sacrificed if sensitivity-compromised sampling (to the optimal $t_{\max}$), instead of uniform truncated sampling, is employed. Exponentially biased sampling is routinely used in conjunction with MDD (Orekhov et al 2003), MaxEnt (Rovnyak et al 2004a), or z.a. FT (Kazimierczuk et al 2006b). For non-linear methods employing *inverse theory* approach (Sec. 2.4.1) a reasonable exponential bias (such as for matched sampling) does not lead to line broadening (Stern et al 2002; Rovnyak et al 2011). In the case of z. a. FT, however, a non-uniform sampling probability density influences lineshapes as predicted by convolution theorem (Kazimierczuk et al 2007).

2.2.5. Applicability of NUS

Finally, we find it useful to indicate practical cases for application of NUS:

1. exhaustive sampling of signals of slowly relaxing coherences (such as TROSY $N^{\pm}H^{\beta}$), particularly at high fields (Rovnyak et al 2004b),
2. sampling of signals of nuclei that are *not* modulated by significant homonuclear couplings (e.g. ^{13}CO, ^{15}N),
3. resolution-limited applications such as coupling measurements (Kazimierczuk et al 2008b), or precise measurement of peak intensity in auto- or cross-correlated relaxation measurements (Stanek et al 2013c),
4. precise chemical shift determination of nuclei exhibiting limited dispersion, e.g. for sequential assignment of IDPs (Zawadzka-Kazimierczuk et al 2012) or non-coding RNAs (Stanek et al 2013b),
5. if reliable connectivities between nuclei can be established by sacrificing some of reasonably good sensitivity. E.g., a quite sensitive HCCH-TOCSY performed as 4D experiment is superior to a pair of 3D (H)CCH- and 3D H(C)CH-TOCSY spectra,
6. decreasing ambiguity of NOE peak assignment in large systems (Tugarinov et al 2005a),
7. fast acquisition of spectra of evolving systems (see Sec. 2.3 for examples) or time-resolved NMR (Mayzel et al 2014). In this case obtained line widths are usually far from natural ones, however, spectra are recorded rapidly.

It should be emphasised that if sensitivity is a major concern, e. g. due to low concentration, the use of NUS to improve resolution is pointless. Except for specific case of *constant-time* evolution, there is always a contradiction between sensitivity and resolution.

2.3. „Fast" acquisition methods unrelated to sampling

2.3.1. Longitudinal relaxation enhancement

Longitudinal relaxation rate (T_1) is commonly underestimated factor that significantly affects sensitivity of NMR experiments. Signal to noise ratio per unit of experimental time is

$$S/N_t \propto \sqrt{\omega_{\mathrm{rep}}}(1-\exp(-(d_1+t_{\mathrm{acq}}/T_1)) \tag{2.18}$$

where $\omega_{\mathrm{rep}} = (d_1+t_{\mathrm{seq}}+t_{\mathrm{acq}})^{-1}$ is the repetition rate, d_1 and t_{acq} are the interscan delay and acquisition time, respectively, and t_{seq} is the pulse sequence length. It can be easily derived that maximal S/N_t of

$$\max S/N_t = A \times 0.71\,(d_{1,opt})^{-1/2} \approx A \times 0.638\,(T_1)^{-1/2} \tag{2.19}$$

is achieved at time optimal interscan delay of $d_{1,opt} \approx 1.26\,T_1 - d_{\mathrm{seq}}$ (A is the proportionality factor in eq. 2.18). The S/N dependence on repetition time and T_1 is shown in Fig. 2.6. Pervushin et al (2002) proposed to selectively manipulate H^Nspins in the out-and-back protein

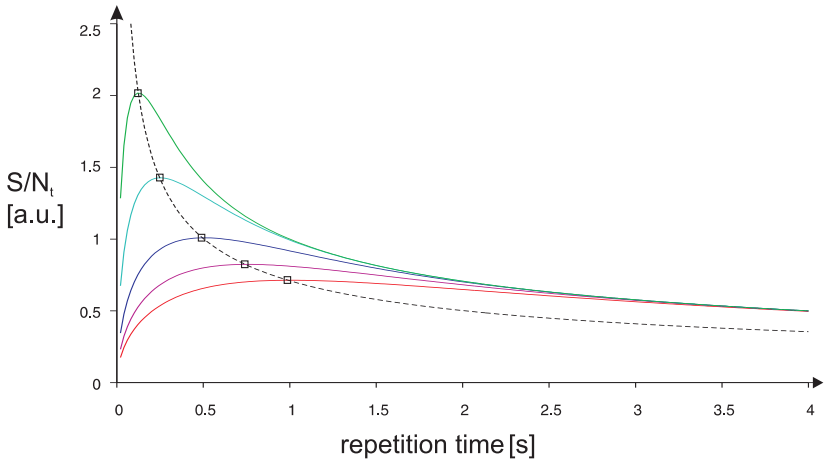


Fig. 2.6 Signal-to-noise per unit time (S/N_t) plotted vs. recovery time for T_1 values of (from *bottom* to *top*) 0.8, 0.6, 0.4, 0.2 and 0.1 s. The values of maximal S/N are indicated with *dashed* line.

experiments (e.g., ^{15}N-TROSY-HSQC, ^{15}N-HSQC, TROSY-HNCA) to avoid saturation of water and ^{13}C-bound ^{1}H spins. It was proven that vast „pool" ^{1}H magnetization kept close to thermal equilibrium (∼80% of all ^{1}H spins) enables fast recovery of perturbed H^N spins due to effective dipolar interactions. Moreover, selective excitation causes that T_1 and also the optimal $d_1 \approx 0.2$ s become fairly independent of polarizing field B_0. Since non-selective T_1 is about 10× larger yet at 600 MHz spectrometers (and increases faster than linearly with B_0), sensitivity gains per unit time as large as 2-3 are achieved. Noteworthy is also that the scatter of T_1 along protein backbone is considerably decreased, thus the optimum d_1 can be determined more precisely than for the non-selective excitation.

The idea was later applied to excitation of imino protons in RNAs (Farjon et al 2009). Despite lower ^{1}H density in RNAs, observed T_1 enhancements are comparable to H^N in proteins. It was proven that crucial here is efficient dipolar relaxation pathway between imino and amino (A, G, U) and H2 (A) protons. In contrast, exhange of imino protons with solvent does not play a crucial role. Additionally, it was demonstated that for non-selective ^{1}H excitation the polarization recovery of the imino protons strongly depends on the amino hydrogen exchange rates ($\omega_{\mathrm{ex}}^{\mathrm{amino}}$), while selective excitation makes imino T_1 fairly independent of $\omega_{\mathrm{ex}}^{\mathrm{amino}}$ and thus more uniform. The applications shown include SOFAST-HMQC in optimal as well as fast pulsing regimes, and ^{1}H-^{15}N TROSY.

It should be clarified here, that significant sensitivity gains of 2-3 are realized only at *fast pulsing regime* of short sequences (e.g., HSQC), and correspond to sensitivity gain relative to non-selective experiments at the same (short) recycling delay. If different recycle delays, optimal for sensitivity of selective and non-selective experiments, are employed in *complicated* pulse schemes, the increase of S/N per unit time is quite moderate (max. 25-30%), if any (Schanda et al 2006). Note, that it is in contradiction to theoretical predictions from eq. 2.19.

Faster T_1-relaxation enables considerable increase in sampling rate (shorter d_1). This has at least two major applications to multidimensional NMR. The first is so-called *real-time* NMR, where experiments are performed in the *fast-pulsing regime*. Most relevant here is *band-Selective Optimized Flip-Angle Short-Transient* (SOFAST) HMQC experiment (Schanda and Brutscher 2005) which combines advantages of selective H^N excitation and optimal Ernst angle (β_{opt}) excitation (Ernst et al 1987),

$$\beta_{\mathrm{opt}} = \arccos(\exp(-T_{\mathrm{rec}}/T_1)) \tag{2.20}$$

The sensitivity gains, determined using eq.

$$S/N \propto \frac{(1-\exp(-T_{\mathrm{rec}}/T_1))\sin(\beta_{\mathrm{opt}})}{1-\exp(-T_{\mathrm{rec}}/T_1)\cos(\beta_{\mathrm{opt}})}\frac{1}{\sqrt{n\,T_{\mathrm{scan}}}} \tag{2.21}$$

where $\beta = \alpha - \pi$, α is the excitation pulse angle, can be as large as one order of magnitude. It was shown that inter-scan delay can be virtually eliminated ($t_{rec} = 1$ ms), leading to $T_{\mathrm{scan}} \approx 100$ ms. This allows to record low-resolution 2D ^{15}N-H^N correlation spectra within few seconds, which is crucial for investigation of kinetics of solvent exchange (Schanda and Brutscher 2005), protein folding (Schanda et al 2007) or ligand binding. SOFAST-HMQC was also employed for selective methyl ^{1}H-^{13}C correlation, however, T_1-enhacements are less pronounced here (Schanda et al 2005). The apparent limitation of SOFAST is that only few pulses and low-power decoupling can be employed to avoid excessive power being deposited in the probe circuits. Effectively, it does not represent a general solution for fast sampling in multidimensional NMR (3D and more).

The second relevant application of T_1-enhancement is *Band-selective Excitation Short-Transient* (BEST) approach to record standard multidimensional experiments, e.g. HNCO or HNCA, in a considerably reduced time (Schanda et al 2006). It was shown that carefully modified multi-pulse sequences are feasible to retain ∼60% of equilibrium unused magnetization at steady-state conditions. Despite somewhat higher scan times, 3D spectra can be recorded using uniform sampling in as few as 10 minutes. Further applications of BEST to sequence-specific protein assignment experiments were demonstated (Lescop et al 2007).

Notable is the combination of TROSY and BEST approaches since absence of heteronuclear decoupling allows to significantly reduce recycle delay without risk of probe damage (Lescop et al 2010). Recently, Favier and Brutscher (2011) showed that additional (non-equilibrium) ^{15}N polarization is created in TROSY experiments under BEST conditions ($T_1^{\mathrm{Hsel}} \ll T_1^{\mathrm{N}}$) from H_z polarization which arises due to ^{1}H T_1 relaxation *during the sequence*:

$$\cdots \xrightarrow{\text{relaxation}} \mathrm{H_z} \xrightarrow{ST2-PT} -\mathrm{N_z}$$

This ^{15}N magnetization can be utilized in the subsequent scan for further enhancement of the TROSY pathway. It is also worth emphasising that both SOFAST and BEST-type experiments feature reasonably good water suppression, which is a significant issue, e.g. for single-scan approach (next section).

Recently, joint use of 3D BEST-TROSY experiments and SOFAST-HMQC enabled NMR characterization of a transient folding intermediate at atomic resolution within few tens of

minutes (Rennella et al 2012). Further applications include in-cell NMR, where monitored transiently populated biomolecules have similar lifetime.

To a certain extent, BEST approach alleviates sampling limitation by increased sampling rate. However, further acceleration is feasible only in combination with *non-uniform sampling*, either radial (Atreya and Szyperski 2004), or random NUS. Although BEST has limited capabilities of enhancing overall sensitivity per unit time, faster pulsing increases number of samples per unit time, which supports spectral reconstruction algorithms. In this context, noteworthy is the application BEST-TROSY experiments in *targeted acquisition* (TA) approach that makes use of both NUS and MDD processing (Jaravine and Orekhov 2006).

In conclusion, T_1-relaxation enhancement enables more efficient use of high-field spectrometers, and provides unique possibilities for severly time-limited studies of evolving systems. Nevertheless, the applicability of the approach is limited to out-and-back experiments, where spin state of different protons can be reliably controlled. Additionally, power delivered to probe per unit time further limits achievable speed-up gains unless decoupling-free (e.g. TROSY) schemes are used. Thus, this approach alone cannot be considered a universal remedy for time-resolution dependence.

2.3.2. Single-scan approach

The idea of single-scan acquisition of multidimensional spectra originates from magnetic resonance imaging (MRI). It was argued by Frydman et al (2002) parametric evolution in the indirect dimension, t_1, resembles a very time-inefficient continuous wave approach. The latter was substituted by FT for 1D spectroscopy to collect signal responses from all signals in the sample *at the same time*. The principle of single scan is to parallelise a known scheme of parametric t_1 evolution

$$\text{preparation} - \text{evolution } (t_1) - \text{mixing} - \text{real-time acquition}(t_2)$$

so that different slices of a sample simultaneously contribute to the acquired signal with different evolution times, t_1. The approach of Frydman generally assumes that a single slice provides sufficient S/N, which is major limitation of the method. Artificial partitioning of sample is introduced using possibly linear pulsed field z-gradients (PFGs). In order to assign different evolution time to spins at different sample z-position, a selective excitation (for real-time) or refocusing (for constant-time evolution) pulses with gradually incremented frequency offset is employed as shown in Fig. 2.7*a*. Since a pair of time-equivalent gradients of amplitude $\pm G_{\text{ex}}$ is applied during excitation, the net evolution is due to chemical shift only. Effectively, every slice undergoes evolution during a different time and contributes to the signal (Fig. 2.7*b*). Frydman et al (2002) proposed to use echo planar spectroscopic imaging (EPSI) to separate resulting slice responses during acquistion time. For this purpose a pair of decoding gradients $\pm G_a$ are used during detection (Fig. 2.7*a*). At this time all coherences acquire a phase $\phi = k\,z + \Omega\,t_2$ where $k = \int_0^{t_2} \gamma_a G_a(t) dt$ varies linearly with t_2 in a „zig-zag" trajectory as shown in Fig. 2.7*c*. The signal is sampled at a very high rate with increments Δ_{acq}, which are no longer time increments for t_2. To understand how the spectral information is obtained, it is convenient to assume homogeneous (uniform) spatial distribution of spins in a sample, $\rho(z) = \rho$. In this case the acquired time-domain signal is (Frydman et al 2003)

$$S(t_1, t_2) = \rho \int_{v_1} \int_{v_2} I(v_1, v_2) \exp(iv_1 t_1) \exp(iv_2 t_2) \mathrm{d}v_1 \mathrm{d}v_2 \tag{2.22}$$

In the above equation it is crucial to note that t_1 dependence is introduced via contributions from different slices, i.e. integration over sample volume. Thus it is z-dependent, e.g. $t_1 =$

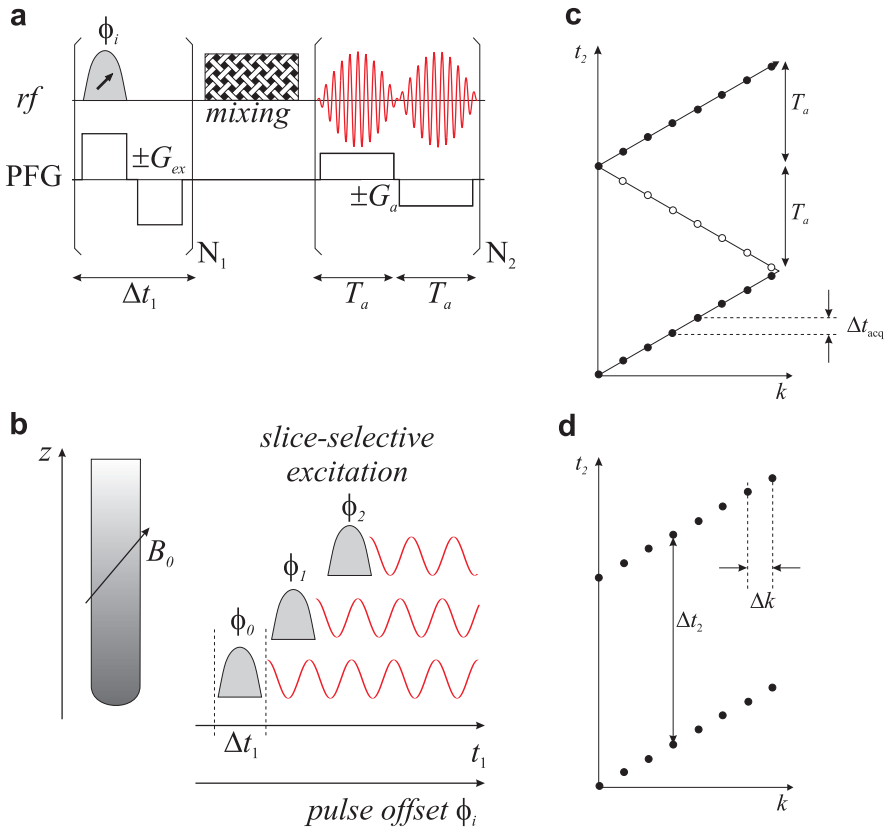


Fig. 2.7 (a) A general scheme of single-scan experiments comprising N_1 slice-selective elements, arbitrarily chosen mixing part, and EPSI detection scheme. (b) The idea of selective excitation. (c) Real-time sampling depitcted as points in (k, t_2) space. (d) Explanation of final data sorting for extraction of FID for conventional FT in t_2. Note that dwell-time, Δt_{acq} is no longer the time increment for t_2, and $\Delta t_2 = 2\,T_a$.

$C(z - z_0)$. For a particular value of k, the data can be re-arranged as shown in Fig. 2.7d to perform FT to convert t_2 into v_2 dependence. Noteworthy, spectral width in the detection dimension is $\mathrm{sw}_2 = (\Delta t_2)^{-1} = (2T_a)^{-1}$, and max. evolution time $t_{\mathrm{max},2} = N_2 \times 2\,T_a$ limited by relaxation as usual. Quite surprisingly, there is no need to perform FT in the indirect dimension. In fact, the signal is collected as

$$S(k, t_2) = \rho \int_{v_1} \int_{v_2} \int_{z} I(v_1, v_2) \exp(iv_1 C(z - z_0)) \exp(iv_2 t_2) \exp(ikz) \mathrm{d}v_1 \mathrm{d}v_2 \mathrm{d}z \qquad (2.23)$$

i.e. in the k-space. There is little point in performing inverse FT to convert it to z-domain, which is equivalent to t_1 domain, and subsequently applying FT to end in v_1 domain. In other words, k space is equivalent to Fourier domain and simple proportionality between v_1 and k exists: $k = -Cv_1$ (Frydman et al 2003). Thus, processing of single-can data involves only FT in direct dimension.

For a complete description, let us mention that proportionality constant C relates to experimental parameters: Δt_1, twice the selective excitation pulse duration (T_p), $\Delta\phi_i$, the phase increment of this pulse, G_{ex}, gradient strength during excitation, and gyromagnetic ration of nuclei in the involved coherence γ_{ex}:

$$C = \Delta t_1 \left(\frac{\Delta\phi_i}{\gamma_{\mathrm{ex}} G_{\mathrm{ex}}} \right)^{-1} \qquad (2.24)$$

Also, the spectral window in the indirect dimension is determined as

$$\mathrm{sw}_1 = \left| \frac{k_{\mathrm{max}}}{C} \right| = \left| \frac{\gamma_a G_a T_a}{C} \right| = \left| \Delta\phi_i \frac{\gamma_a G_a T_a}{2T_p \gamma_{\mathrm{ex}} G_{\mathrm{ex}}} \right| \qquad (2.25)$$

where a subscripts refer to coherence, and gradient strength and duration during *acquisition*. It can be inferred from above eq. that spectral width is limited by available gradient strengths of G_a and G_{ex}, and their capability for fast switching. In contrast, the resolution in k-space (also v_1) is determined by physical dwell time during acquisition, Δt_{acq}. Noteworthy, sampling rate must be increased by a factor of $2N_k = 2T_a/\Delta t_{\mathrm{acq}}$ with respect to conventional acquisition to provide digitization of N_k points in k-space. This, however, requires filter bandwidths to be increased leading to $\sqrt{2N_k}$ amplification of incoming noise (Frydman et al 2003). In effect, S/N in a single-scan experiment is considerably lower than for a standard acquisition and may be considered a price for fast data collection.

Single-scan represents a very general concept, and various mixing element can be used to perform virtually all kinds of correlation spectroscopy (Frydman et al 2003). Also multidimensional (3D, 4D) extensions were proposed (Shrot and Frydman 2003). Single scan is most successful in organic chemistry, or high-throughput applications, reaction monitoring (Gal et al 2006), liquid chromatography (Shapira et al 2004) or when combined with hyperpolarization (DNP) (Giraudeau et al 2009). Alternatively, it could be regarded as alternative to sparse sampling for ultra-fast acquistion of very sensitive spectra (e.g., HSQC) of evolving biological systems (Lee et al 2010). Noteworthy is also joint use with SOFAST-HMQC (Gal et al 2007).

Unfortunately, the applicability of single-scan NMR to protein samples is currently limited by its inherently low sensitivity and demanding spectrometer hardware requirements. Additionally, H_2O suppression poses a great challenge, making it quite troublesome for routine use in biomolecular NMR.

Despite inherent sensitivity limitation, there are recently initiatives between our and Frydman's group to combine spatial encoding with NUS. The aim is to reduce dimensionality of sparsely sampled domain so it could be processed using high fidelity algorithms such as Compressed Sensing.

2.3.3. Multiple receivers and other ideas

Multidimensional NMR is completely dominated by ^{1}H-detected experiments due to its intrinsically high sensitivity. However, ^{13}C-detection has also gained considerable attention in biomolecular NMR over last decade, owing to unique features such as absence of solvent (H_2O) signal, generally better signal dispersion in wide spectral window, insusceptibility to chemical exchange, access to proline residues, and low number homonuclear couplings, which are also quite well-defined (Bertini et al 2004; Bermel et al 2006, 2012). With the advent of multiple receivers built-in to modern NMR hardware it has become possible to devise new experimental scheme which are capable of simultaneous acquisition of NMR signals of different nuclei (Kupče et al 2006).

Initial applications of this new approach were in the NMR of organic species. *Parallel Acquisition NMR, an All-in-one Combination of Experimental Applications* (PANACEA) which combines INADQUATE, HSQC and HMBC allows to solve the topology of simple organic compounds in a single experiment (Kupče and Freeman 2008). Biomolecular NMR application are obviously more demanding, and non-trivial changes in pulse sequence design must be made. One approach was to perform detection of ^{13}C (CO) signal during the sequence, and utilize remaining coherent signal for a more sensitive ^{1}H-detected part. The example of joint 2D (HA)CACO and 3D (HA)CA(CO)NNH was shown which exploits the following

coherence pathways (Kupče et al 2010):

$$^1\mathrm{H}\alpha \xrightarrow{^1J_{\mathrm{CH}}} {}^{13}\mathrm{C}\alpha\ (t_1) \xrightarrow{^1J_{\mathrm{C}\alpha\mathrm{CO}}} {}^{13}\mathrm{CO}\ (t_2)$$
$$^1\mathrm{H}\alpha \xrightarrow{^1J_{\mathrm{CH}}} {}^{13}\mathrm{C}\alpha\ (t_1) \xrightarrow{^1J_{\mathrm{C}\alpha\mathrm{CO}}} {}^{13}\mathrm{CO} \xrightarrow{^1J_{\mathrm{CON}}} {}^{15}\mathrm{N}\ (t_2) \xrightarrow{^1J_{\mathrm{HN}}} {}^1\mathrm{H}^{\mathrm{N}}\ (t_3)$$

A more recent approach alleviates some of sensitivity and resolution issues by performing parallel acquisition *at once*, at the end of the pulse sequence rather than sequentially (Kupče and Kay 2012). This requires the coherence transfer pathways to split during the sequence. E.g., it was shown that independent manipulation of $\mathrm{N_z}$ and $\mathrm{N_zCO_z}$ terms can lead to parallel acquistion of 2D HSQC and 2D NCO spectra (Kupče and Kay 2012). Moreover, this methodology can further be exploited to record 3D ^{1}H-detected HNCA || 3D ^{13}CO-detected CANCO.

Recently, Reddy and Hosur (2013) proposed a novel scheme for parallel detection of 3D HA(CA)NH and 3D HACACO. The coherence pathways are slightly different from that proposed by Kupče et al (2010):

$$^1\mathrm{H}\alpha\ (t_1) \xrightarrow{^1J_{\mathrm{CH}}} {}^{13}\mathrm{C}\alpha\ (t_2) \xrightarrow{^1J_{\mathrm{C}\alpha\mathrm{CO}}} {}^{13}\mathrm{CO}\ (t_3)$$
$$^1\mathrm{H}\alpha \xrightarrow{^1J_{\mathrm{CH}}} {}^{13}\mathrm{C}\alpha\ (t_1) \xrightarrow{^1J_{\mathrm{C}\alpha\mathrm{N}},\,^2J_{\mathrm{C}\alpha\mathrm{N}}} {}^{15}\mathrm{N}\,(i, i{-}1)\ (t_2) \xrightarrow{^1J_{\mathrm{HN}}} {}^1\mathrm{H}^{\mathrm{N}}\ (t_3)$$

and are both detected simultaneously at the end of the sequence.

Time-shared NOESY experiments represent a similar approach to improve time-efficiency of NMR spectroscopy. Various schemes were proposed to realize independent magnetization transfer pathways, e.g. for ^{15}N- and ^{13}C-bound ^{1}H spins (Pascal et al 1994; Farmer and Mueller 1994; Sattler et al 1995; Xia et al 2001, 2003; Frueh et al 2006a; Coggins et al 2012). These methods utilize extended phase cycles to separate ^{15}N-, ^{13}C(aro)-, and ^{13}C(ali)-edited NOESY spectra. While the goal of simultaneous acquisition of various NOESY signals is obtained, some concomitant compromises with respect to sensitivity and resolution are made.

A quite distinct approach was presented by Diercks and Orekhov (2005), who proposed intermediate storage of *anti*-TROSY polarization in a so-called queued-TROSY (qTROSY) scheme:

$$^1\mathrm{H}^{\mathrm{N}} \xrightarrow{\text{INEPT}} \xrightarrow{\mathrm{S^3E}} \begin{cases} \mathrm{N}^{\pm}(\frac{1}{2} + \mathrm{H_z}) \longrightarrow \ldots \xrightarrow{\text{ST2-PT}(+|-)} \mathrm{H}^{\mp|\pm}(\frac{1}{2} + \mathrm{N_z})\ \text{scan A} \\ \mathrm{N_z}(\frac{1}{2} - \mathrm{H_z}) \longrightarrow \ldots \xrightarrow{\text{ST2-PT}(+|-)} \pm\,\mathrm{H_z}(\frac{1}{2} - \mathrm{N_z})\ \text{stored} \end{cases}$$

The stored component is utilized immediately after acquisition of the first signal, thus avoiding reequilibration delay between two acquisitions

$$\mathrm{H_z}(\frac{1}{2} - \mathrm{N_z}) \xrightarrow{90^\circ_y(^1\mathrm{H})} \xrightarrow{\text{reversed ST2-PT}} -\,\mathrm{N_y}(\frac{1}{2} + \mathrm{H_z}) \longrightarrow \ldots \xrightarrow{\text{ST2-PT}} \text{scan B}$$

While this generally does not allow for improved sensitivity due to relaxation and other source of signal loss, two arbitrarily different $\mathrm{H^N}$-start experiments can be performed in a roughly the same time. Therefore, the best way to utilize qTROSY is to combine two experiments of intrinsically different sensitivity, recording the more sensitive one as scan B (less optimal).

In conclusion, all above mentioned approaches to accelerated acquisition offer specific advantages, however, attainable time-savings are quite limited, or compromise resolution. They are not as general as NUS, and should support rather than replace it.

2.4. Spectral reconstruction methods for randomly NUS data

2.4.1. Spectral reconstruction as an inverse problem

The problem of spectral reconstruction may be formulated as follows: find the spectrum S which, transferred to time-domain using inverse Fourier Transform, is in agreement with measured signal d:

$$\mathbf{F}\mathbf{T}^{-1}(S) = d \tag{2.26}$$

For the case of discrete data ($d_k = d(t_k)$), the problem can be expressed as a set of K linear equations

$$\sum_{m=1}^{M} \mathbf{F}_{km} S_m = d_k, \quad \text{for} \quad k = 1, \ldots, K \tag{2.27}$$

where elements of the *inverse* Fourier matrix are $\mathbf{F}_{km} = \frac{1}{M} \exp(i\,\omega_m t_k)$, and M equidistant spectral points are $S_m = S(\omega_m) = S(m\Delta_\omega), \Delta_\omega = sw\,/M$[1]. Let me comment that for linearly sampled data ($t_k = k\Delta_t$, $k = 0, \ldots, K-1$), and equal number of data and spectral points ($K = M$), the matrix $\mathbf{F}$ is unitary[2], i.e. $\mathbf{F}^\dagger\mathbf{F} = 1$ (where $\dagger$ denotes Hermitian conjugate). In this special case the solution is trivial

$$S = \mathbf{F}^{-1} d = \mathbf{F}^\dagger d \tag{2.28}$$

and plainly corresponds to conventional discrete Fourier transformation.

The form of Eq. 2.27 enables generalization to the case of incomplete data, i.e. $K \ll M$. Obviously, the Fourier matrix is rank-deficient as there are more unknowns than equations. Consequently, there is an infinite number of possible solutions. One has to introduce specific assumptions to obtain an *unique* solution. These assumptions represent our knowledge (or belief) about the supposedly correct „physical" solution. The reconstruction problem is converted into a problem of minimization of specific *regularizing* functional $\mathcal{R}(S)$

$$S : min\ \mathcal{R}(S) \tag{2.29}$$

subject to constraint that the solution agrees with experimental data d within assumed limits, usually determined by experimental errors (compare to eq. 2.27)

$$\sum_{k=1}^{K} \left(\sum_{m=1}^{M} (\mathbf{F}_{km} S_m) - d_k \right)^2 \leq C_0 \tag{2.30}$$

The constrained problem can be converted into unconstrained optimization by introduction of Lagrange multiplier λ. The resulting objective function is

$$\mathcal{Q}_\lambda(S) = \mathcal{R}(S) + \lambda \mathcal{C}(S) \tag{2.31}$$

Parameter λ ($0 < \lambda < \infty$) reflects the balance between the agreement with experimental data (constraint term $\mathcal{C}$) and our prior regarding desired appearance of spectrum (regularizing term $\mathcal{R}$). Although there is no one correct value of λ, the reasonable choice should result in a probable value of constraint term. Since the latter is closely associated with χ^2 statistic,

$$\chi^2 = \sum_{k=1}^{K} \left(\frac{s_k - d_k}{\sigma_k} \right)^2 \tag{2.32}$$

[1] Any rotation of spectral points is allowed as predicted by aliasing theorem. E.g. $\omega_m = m\Delta_\omega - 1/2\,sw$

[2] if normalization factor $1/M$ is replaced by $1/\sqrt{M}$

one may expect that probable values of restraint term are

$$\mathcal{C}(S) \approx \sigma^2(K \pm \sqrt{K}) \tag{2.33}$$

where we assume that all experimental values have the same variance σ^2. Significantly smaller values indicate that found solution matches experimental data unlikely well. In contrast, greater values than predicted by Eq. 2.33 suggest that agreement with experimental data is less favoured than assumed properties of solution. It is essential to determine the realistic error of measured data (σ^2). In practice, somewhat empirical, or iterative adjustment of λ is necessary to obtain desired agreement with experimental data.

The final solution of the ill-posed initial problem (Eq. 2.27) is of course dependent on the form of regularizing functional $\mathcal{R}$. In the following sections, a number of ideas for spectral reconstruction are discussed. They are not given in the chronological order, but rather to emphasise relationships and differences between them.

2.4.2. l_2-norm reconstruction and „non-uniform" FT

One of the simplest forms of a *regularizing* functional is a positive-definite ($S\mathbf{R}S > 0$ for every vector $S \neq 0$), nondegenerate quadratic form, which has a simple matrix representation:

$$\mathcal{R}(S) = S\,\mathbf{R}\,S \tag{2.34}$$

This defines the class of *linear* regularizing methods (Press et al 2007, §19.5). Let us note that in this case the objective function of the form

$$\mathcal{Q}_\lambda(S) = S\,\mathbf{R}\,S + \lambda(\sigma^2\,\chi^2(S)) \tag{2.35}$$

consists of two positive-definite quadratic forms. The latter one is degenerate, however, the regularizing term guaranteers non-degeneracy of their sum. Consequently, the overall problem is well-defined and unique solution exists. It can be found as follows:

$$\begin{aligned} 0 &= \nabla Q = \nabla(S\,\mathbf{R}\,S + \lambda|\mathbf{F}S - d|^2) \\ 0 &= 2\,\mathbf{R}\,S + 2\lambda\mathbf{F}^T(\mathbf{F}S - d) \\ 0 &= (\mathbf{R} + \lambda\mathbf{F}^T\mathbf{F})S - \mathbf{F}^T d \end{aligned} \tag{2.36}$$

We immediately recognize that S is the solution of linear $M \times M$ (algebraic) problem, and can be solved using standard methods of linear algebra, e. g. LU-decomposition (Press et al 2007, §2.3). One of the most relevant cases is the so-called Tikhonov regularization, which assumes *smoothness* of solution, or, minimal norm of its first derivative. This leads to the tridiagonal form of the matrix $\mathbf{R}$. The more detailed discussion is beyond the scope of this thesis, and in the following we will restrict ourselves to the simplest, although also quite inoptimal, choice of *regularizing* functional: $\mathbf{R} = \mathbf{1}$. It is commonly referred to as *zeroth-order regularization* (Press et al 2007, §19.4.1), or l_2-norm ($\mathcal{L}_2(S) = |S|^2$) reconstruction:

$$\mathcal{Q}_\lambda(S) = S\,\mathbf{1}\,S + \lambda|\mathbf{F}S - d|^2 = |S|^2 + \lambda\sigma^2\,\chi^2(S) \tag{2.37}$$

One interesting observation is that the initial ill-posed (degenerate) linear problem

$$\boldsymbol{F}S = d \tag{2.38}$$

is equivalent to minimization of functional (2.37) in the limit of $\lambda \to \infty$. The standard solution of this problem, featuring minimal power (l_2-norm), can be found employing *singluar value decomposition* (Press et al 2007, §2.6)

$$\mathbb{F} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{2.39}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $K \times K$ and $M \times M$ unitary matrices, and $\mathbf{W}$ is diagonal matrix containing K non-zero singular values. This possibility was extensively exploited in the Author's M.Sc. thesis. However, in the case of sparse *on-grid* sampling, the matrix $\mathbf{F}$ has fairly trivial singular values, $\mathbf{W} = \mathbf{1}$, and the solution can be found in a more straightforward way as explained below.

Let us recall the Parseval's theorem (Sundararajan 2001, §4.11), which relates power of time and spectral domain representations:

$$\sum_{k=0}^{N-1} |s_k|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |S_n|^2 \tag{2.40}$$

It follows immediately that minimal l_2-norm spectral reconstruction also corresponds to the time-domain signal of minimal power. This provides insight into the nature of l_2-norm reconstruction: it simply replaces unmeasured signal samples with zeros since it does not increase the power of signal. Hence, the method is sometimes referred to as *zero-augmented Fourier Transform* as suggested by Professor J.C. Hoch (University of Connecticut). It is not necessary to find pseudoinverse of matrix $\mathbf{F}$ using SVD, since it has the explicit form of ${\mathbf{F}^*}_{mk} = \exp(-i\,\omega_m\,t_k)$ (for on-grid sampling only). Spectrum estimate can be thus evaluated using formula

$$S_m = \sum_{t_k \in \Omega} s_k \exp(-i\,\omega_m\,t_k) \tag{2.41}$$

which is plainly discrete FT, except that only measured samples, belonging to the sample schedule Ω, are included in the sum. Apparent similarity to DFT has lead to emergence of a quite vague term of „*non-uniform FT*" (nuFT or nuDFT), which we prefer to avoid.

One should emphasise that both pseudoinverse or eq. 2.41 provide exact solution of eq. 2.39, which is extremely unlikely in terms of χ^2 statistic. On the other hand, zero-filling does not reduce or introduce new information to the signal, and removal of λ parameter makes it a useful, non-parametric method for spectral estimation.

In addition, a quite unexpected relationship to Monte Carlo integration (MCI) can be deduced from eq. 2.41. Let us recall the basic theorem of MCI (Press et al 2007, §7.7):

$$\int f \mathrm{d}V \approx V\langle f \rangle \pm V\sqrt{\frac{1}{K}(\langle f^2 \rangle - \langle f \rangle^2)} \tag{2.42}$$

which states that integral of function f can be approximated, within known limits, to the mean of its values at randomly uniformly chosen points x_i:

$$\langle f \rangle = \frac{1}{K} \sum_{k=1}^{K} f(x_i)$$

(times volume V). We thus interpret eq. 2.41 as an approximation of the Fourier integral

$$S(\omega) = \int_0^{t_{max}} s(t) \exp(-i\,\omega\,t) \mathrm{d}t \tag{2.43}$$

using a limited number of K samples. In fact, MCI is usually employed for multidimensional integrals of functions which evaluation requires non-negligible computational time, and thus the number of function samples is severly limited. This is exacltly the difficulty encountered in multidimensional NMR, where each indirectly sampled point requires at least a few seconds!

The close relationship between Fourier sums and estimation of Fourier integrals allows us to propagate known features of FT, namely *linearity*, *convolution*, and various *symmetry* properties (Kazimierczuk et al 2012). Additionally, non-uniform distribution of samples according to probability density $p(t)$ (weighting probability approach, WP) appears equivalent to estimation of Fourier integral of a modified signal:

$$\sum_{t_k \in \Omega} s_k \exp(-i\,\omega_k\,t_k) \propto \int_0^{t_{max}} s(t)p(t)\exp(-i\,\omega t)\mathrm{d}t$$

where we recognize $p(t)$ playing role of *apodization* function (weighted samples approach, WS) (Kazimierczuk et al 2007).

Kazimierczuk et al (2009) considered properties of approximated Fourier integrals using the notion of *point distribution function* (PDF), which is the real part (cosine transform) of FT of the constant signal, $s(t) = 1$:

$$S(\omega) = \int_0^T \cos(\omega t)\mathrm{d}t$$

It was derived that its estimation by Fourier sum (2.41) has a probability density

$$h(y) = \frac{1}{\pi\sqrt{1-y^2}} \tag{2.44}$$

where $y = \cos(\omega t)$ is a random variable for a given frequency $\omega \neq 0$ and uniformly distributed random variable t. Because of the symmetry the mean value vanishes:

$$\langle y \rangle = \int_{-1}^{1} \frac{y}{\pi\sqrt{1-y^2}}\mathrm{d}y = 0$$

while its variance, equal to $\langle y^2 \rangle$, is

$$\mathrm{Var}(y) = \langle y^2 \rangle = \int_{-1}^{1} \frac{y^2}{\pi\sqrt{1-y^2}}\,\mathrm{d}y = 2\int_0^1 \frac{y^2}{\pi\sqrt{1-y^2}}\,\mathrm{d}y = \frac{2}{\pi}\int_0^{\pi/2} \sin^2 t\,\mathrm{d}t = \frac{1}{2}$$

It is essential that all terms in the sum (2.41) for independent random variables t_k have the same probability density 2.44. According to Central Limit theorem, in the limit of large number of variables K the average of y_i features a Gaussian distribution with variance

$$\mathrm{Var}(\frac{1}{K}\sum_i y_i) = \frac{\mathrm{Var}(y)}{K} = \frac{1}{2K} \tag{2.45}$$

irrespectively of the distribution of variables y_i. Let us also note, that for $\omega = 0$ the Fourier sum is exact, of zero variance, and equals to $S(0) = K$. Since we can express signal-to-artefact (S/A) ratio as

$$S/A = \frac{S(0)}{\sqrt{\mathrm{Var}(S(\omega))}} \tag{2.46}$$

and the variance in the denominator relates to eq. 2.45 as

$$\mathrm{Var}(S(\omega)) = \mathrm{Var}(K\frac{1}{K}S(\omega)) = K^2\mathrm{Var}(\frac{1}{K}S(\omega)) = K^2\frac{1}{2K} = \frac{K}{2} \tag{2.47}$$

the value of S/A is

$$S/A = \frac{K}{\sqrt{\frac{K}{2}}} = \sqrt{2K} \tag{2.48}$$

This formally proves that for randomly sampled signal the relative intensity of artefacts depends entirely on the number of samples[3], and other factors such as relative sampling density are irrelevant. This conclusion is valid for unrestricted (off-grid) random sampling of uniform probability distribution, and also for on-grid sampling at low sampling density (see sec. 2.2.2).

This fact, also noted earlier empirically in simulations (Kazimierczuk et al 2006b, 2007), has profound consequences. Firstly, NUS artefacts exhibit similar behaviour to thermal noise, and their relative magnitude does not change upon accumulation of signal samples. Experiments which provide signal of S/N in a single sample $\gg \sqrt{2}$ are *sampling-limited*, while those of S/N $\ll \sqrt{2}$ are *sensitivity-limited*. In the latter case, the influence of sampling artefacts can be neglected. Noteworthy, it can be quickly determined which out two classes a specific experiment belongs to.

Secondly, one should comment that suppression of sampling artefacts by increasing number of samples is quite inefficient. The convergence of $\sqrt{K}^{-1}$ is fairly slow, and for sampling-limited experiments more advanced processing schemes have to be employed.

Thirdly, since there is no coupling between sampled volume (t_{max}), spectral width (sw) and intensity of artefacts, one can extend sampling to any reasonable limits, i.e. determined by relaxation of measured signals. This conclusion also applies to sampling in a multidimensional space (Kazimierczuk et al 2009), and indeed, triggered development of numerous high-dimensional, maximally-resolved protein NMR experiments (Zawadzka-Kazimierczuk et al 2010, 2012; Motáčková et al 2010; Kazimierczuk et al 2009, 2010; Bermel et al 2012; Nováček et al 2011, 2012, 2013; Wen et al 2011).

The analysis of numerical cost of l_2-reconstruction reveals its additional advantages. The direct solution of inverse algebraic $M \times M$ problem with finite λ (eq. 2.36) is of order of M^3 operations. However, usually a series of data vectors are processed for consecutive points in the direct dimension, and only the cost of back-substitution ($M \times K$) is relevant since LU-decomposition is not repeated. Similar conclusions can be drawn for SVD (limiting case of $\lambda \to \infty$). Thus, the numerical cost is the same as for direct Fourier sum, $M \times K$ (eq. 2.41). Moreover, numerical cost can be reduced by a factor of 2 if only real part of the spectrum is to be computed. On the contrary, non-linear methods (discussed below) require iterative solution of the invese problem (eq. 2.31). For gradient optimization methods and functionals close to quadratic forms it requires approx. M iterations (Press et al 2007, §10.8), each including inverse FT (of at least $M \log M$ operations).

Further benefits are evident in a multidimensional indirect spectral domain. In the case of general off-grid sampling, direct Fourier sum requires $M \times K$ operations, where M represents overall number of spectral points

$$M = \prod_{i=1}^{D} M_i$$

Direct methods for solving the inverse linear problem cannot be easily applied since any decomposition of $M \times K$ matrix requires prohibitive storage and computational time. Iterative methods could still be applied, however, they are again quite computationally intensive. Noteworthy, these limitations vanish if samples are arranged on a Cartesian grid as the l_2

[3] In the original work of Kazimierczuk et al (2009) only proportionality was derived. Also, consideration of Monte-Carlo integration error lead us to the same conclusions (Kazimierczuk et al 2012).

reconstruction factorizes into 1D problems. In practice, direct Fourier sums yield the same results without additional complexity. Therefore, direct solution of *inverse problem* should be employed only if stable deconvolution is desired (see Section 2.4.6).

In contrast, non-linear methods quickly encounter computational barrier for high-resolution ND-spectra ($N \geq 3$) as number of variables in minimization, M, grows exponentially with dimensionality D. In fact, this is *implicit* sampling limitation, since $M_i \propto t_{max,i} \times sw_i$. In this context, the extraordinarily small cost of multidimensional Fourier sum explains its originally proposed, a little obsolete and imprecise name, *multidimensional FT (MFT).*

Another striking feature of Fourier sums (eq. 2.41) is that spectrum can be computed locally, i.e. only at arbitrary chosen spectral points. This idea, referred to as *sparse* MFT (Kazimierczuk et al 2009), enabled practical use of high-resolution 5D (and more) spectra and determination of scalar couplings in 3D and 4D spectra (Kazimierczuk et al 2008b, 2010).

Finally, let us comment on some limitations of l_2-norm reconstruction.

1. As we already pointed out, S/A increases slowly with the number of samples, thus the method might be inappropriate for very short NUS experiments. For extremely simple signals containing a single resonance, which might be the case of HNCO of well folded small proteins, a minimal reasonable S/A of 10 is obtained for $K \geq 50$.

2. In addition, if multiple signals are present, their artefacts accumulate. This causes loss of information even if all signals are of similar intensity. Fig. 2.8 shows the effect of spectral crowding on quality of z.a. FT spectra. This demonstrates that additional accumulations are required to recover true spectral information even for relatively simple experiments. Notably, H^N-detected spectra of IDPs feature poor dispersion in the direct dimension, which results in massive accumulation of NUS artefacts in the indirect domain. This causes that these experiments are typically acquired unnecessarily long to achieve reasonable quality. As argued by Wen et al (2011), this overhead can easily be avoided using simple artefact suppression methods.

3. The last troublesome situation where z.a. FT is of limited applicability is a signal containing components of significantly different intensity. As illustrated in Fig. 2.9, a small signal is completely obscured by NUS artefacts originating from stonger resonances. High dynamic range of signal amplitudes is characteristic for important class of NOESY experiments, but is also not uncommon to observe in H^N-detected experiments if some protons undergo fast exchange with the solvent (e.g., in IDPs).

2.4.3. CLEAN

CLEAN, proposed by Högbom (1974), is probably one of the oldest, albeit still useful algorithm for processing of irregularly sampled data. It originates from the field of radioastronomy, where two-dimensional sky brightness maps are synthesised using data from a certain number of interferometers. Practical considerations suggest that ideal equidistant spacing over the Globe is not achievable. Additionally, data are partially corrupted by malfunctioning of hardware, or distorted by Moon occultations, and have to be discarded. CLEAN was developed to cope with both above mentioned difficulties.

As pointed out by Högbom (1974), reconstruction of all details in high-resolution Fourier maps is not feasible from partial data, and CLEAN is only applicable under certain assumptions. The model of *„dark spectra"* is employed, which specifies that spectral power is concentrated in a small number of *„bright"* spots while vast spectral areas are *„empty"*, i.e.

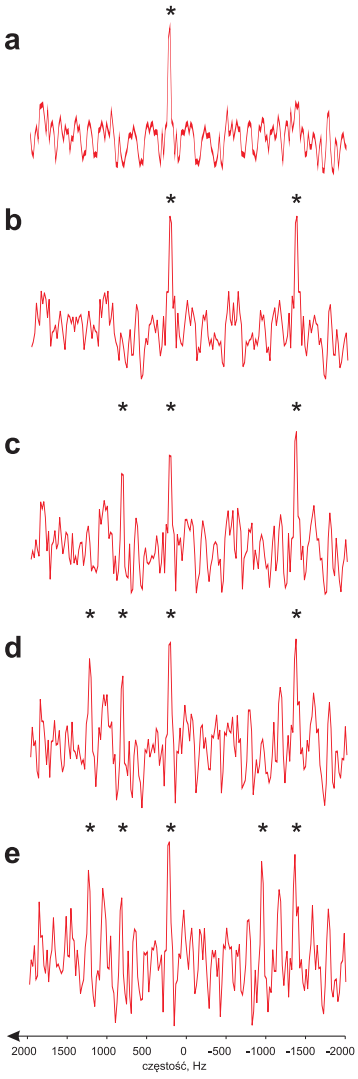


Fig. 2.8 z.-a. FTspectra of signals containing components of *equal* intensity and decay rate. Correct peak positions are marked with stars in every panel (*a*-*e*). Accumulation of NUS artefacts leads to distorted intensities of peaks. Additionally, artefact interferences may be interpreted as *false positives* since they might exceed the intensity of genuine peaks (panel *e*).

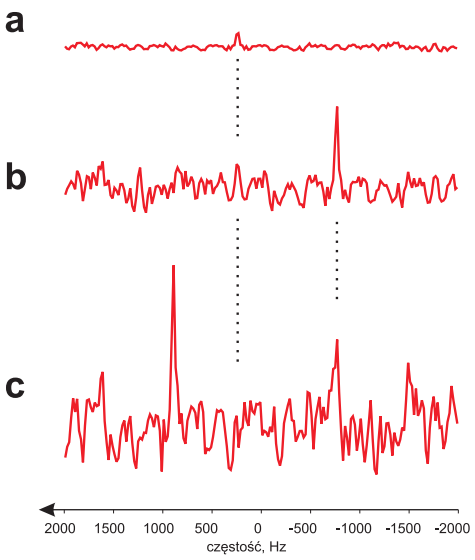


Fig. 2.9 z.a. FT spectra of (*a*) a single resonance of unit intensity, (*b*) with an additional signal of relative intensity of 5, and (*c*) with the third signal of intensity 10. Clearly, smaller peaks are hardly detectable in the presence of stronger sources of NUS artefacts.

dominated by frequency-independent noise. Let us comment, that this model is particularly well-suited to multidimensional NMR spectra.

Contrary to l_2-reconstruction, CLEAN is an iterative algorithm. On entry, z.a. FT spectrum $D(\omega)$ is assumed, which is a convolution of genuine continuous spectrum $S(\omega)$, and FT of sampling function, $\text{Ш}(t)$ (shah):

$$S^{(0)} := D(\omega) = \mathcal{FT}\{\text{Ш}(t) \cdot s(t)\} = S(\omega) * \text{Ш}(\omega) \tag{2.49}$$

The basic idea is that $\text{Ш}(\omega)$ is known, and can be determined as FT of a constant signal sampled at the given points. This function, also referred to as *point spread function* (PSF) or „dirty mask", is to be deconvolved from z.a. FT spectrum to yield „clean" result. Intuitively, one can identify the sources of artefacts by convolution of $D(\omega)$ and PSF. The following is thus interpreted as probability density of artefact source:

$$p(\omega) = (S * \text{Ш})(\omega) \tag{2.50}$$

Convolution theorem immediately brings a quite unexpected result

$$p(\omega) = (S * \text{Ш})(\omega) = \mathcal{FT}\{(s(t) \cdot \text{Ш}(t)) \cdot \text{Ш}(t)\} = \mathcal{FT}\{\text{Ш}(t) \cdot s(t)\} = D(\omega) \tag{2.51}$$

i.e. z.a. FT is itself peak probability. In every iteration, the spectral point of the largest absolute amplitude is selected, $S_{\max} = S^{(i)}(\omega_{\max})$. Subsequently, the spectrum is updated by subtraction of a normalized „dirty mask" shifted to its position, $\omega_{\max}$, and down-scaled by a *loop gain* factor $0 < \gamma \leq 1$:

$$S^{(i+1)} = S^{(i)} - \gamma S_{\max} \text{Ш}(\omega - \omega_{\max}) \tag{2.52}$$

Provided that selected peak was genuine, one obtains a new spectrum with decreased level of artefacts. Extracted „clean" components are accumulated in a separate location, $R(\omega)$, termed „replica"

$$R^{(i+1)} = R^{(i)} + \gamma S_{\max} \delta(\omega - \omega_{\max}); R^{(0)} \equiv 0 \tag{2.53}$$

The process, illustrated in Fig. 2.10, is repeated until no significant spectral features are present in *residual* spectrum $S^{(i_{\max})}$. The termination condition that indicated the fixed point of the algorithm is formulated as follows:

$$|S_{\max}| < \alpha \cdot \sigma_{i_{\max}} \tag{2.54}$$

where α is a fixed arbitrarily chosen parameter (usually integer between 3 and 5), and $\sigma_{i_{\max}}$ is a reliable measure of noise. It is crucial to ensure that this criterion prevents from detection of artificial spectral features, which is particularly troublesome if noise has non-Gaussian and/or frequency-dependent distribution.

The final spectrum is constructed by reintroduction of the residual, $S^{(\max)}$, to the „replica":

$$S_{\text{CLEAN}} := R^{(\max)} + S^{(\max)} \tag{2.55}$$

It might be tempting to discard residual spectrum to retain only strong peaks and apparently improve separation of peaks and noise. However, it was criticized by Shaka et al (1984) as its effect is merely cosmetic: in fact sensitivity is not improved at all, and small peaks present in the residual part may be lost.

It is quite informative to consider CLEAN operation in time domain. Inverse FT of eq. 2.52 yields

$$S^{(i+1)}(t) = S^{(i)}(t) - \gamma S_{\max} \text{Ш}(t) \exp(i\omega_{\max} t) \tag{2.56}$$

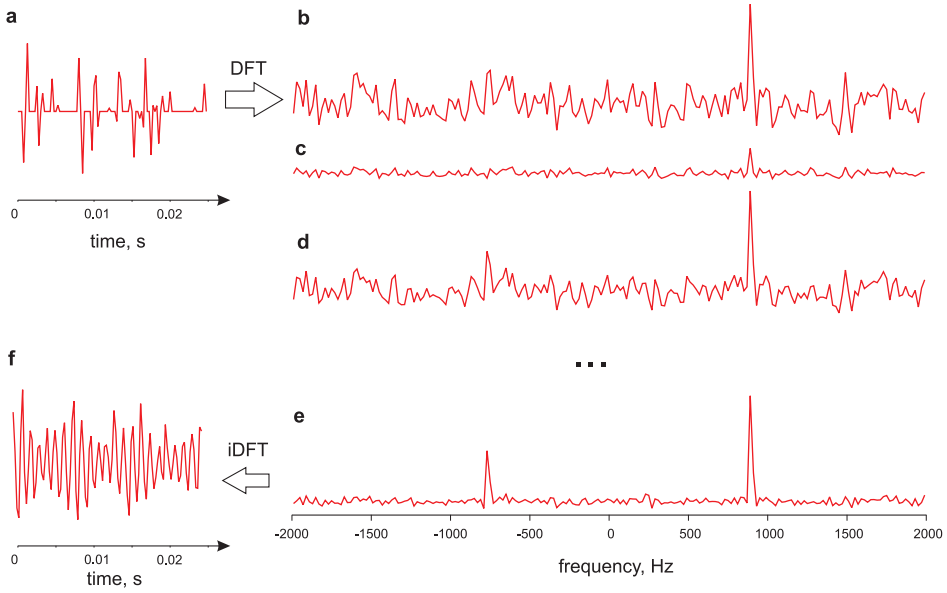


Fig. 2.10 The principle of CLEAN processing demonstrated on a noiseless signal comprising three oscillating components of relative amplitudes 1:5:10 and equal decay rates. The sparsely sampled time-domain signal (*a*) is Fourier transformed with zero-filling of gaps to yield initial spectrum (*b*). Next, *dirty mask* scaled by largest peak intensity and *loop gain* γ (*c*) is subtracted to generate *residual spectrum*. Reconstruction at this stage would yield spectrum *d*. The fixed-point of CLEAN shown in panel *e* corresponds to reconstructed time-domain signal *f*. Note that the smallest peak ($\nu \approx 240$ Hz) is still buried in remaining NUS artefacts.

where we exploited the fact that $\text{Ш}(\omega - \omega_{\max}) = \text{Ш}(\omega) * \delta(\omega - \omega_{\max})$. Similarly, inverse FT of eq. 2.53 leads to

$$R^{(i+1)}(t) = R^{(i)}(t) + \gamma S_{\max} \exp(i\omega_{\max} t) \tag{2.57}$$

Therefore, the final signal, which is also inverse FT of eq. 2.55, can be represented as

$$\begin{aligned} s_{CLEAN}(t) &= \gamma \textstyle\sum_{i=1}^{i_{\max}} (S^i_{\max} \exp(i\omega^i_{\max} t)) + S^0(t) - \gamma \sum_{i=1}^{i_{\max}} (S_{\max} \text{Ш}(t) \exp(i\omega^i_{\max} t)) \\ &= s(t)\text{Ш}(t) + \gamma \textstyle\sum_{i=1}^{i_{\max}} S^i_{\max} \exp(i\omega^i_{\max} t)(1 - \text{Ш}(t)) \end{aligned} \tag{2.58}$$

Let us consider its values at measured points, which can be obtained by side-by-side multiplication by sampling function, $\text{Ш}(t)$:

$$s_{CLEAN}(t)\text{Ш}(t) = s(t)\text{Ш}(t)\text{Ш}(t) = \gamma \sum_{i=1}^{i_{\max}} S^i_{\max} \exp(i\omega^i_{\max} t)(\text{Ш}(t) - \text{Ш}(t)\text{Ш}(t)) \tag{2.59}$$

By utilizing the fact, that $\text{Ш}(t)\text{Ш}(t) = \text{Ш}(t)$, we immediately get

$$s_{CLEAN}(t)\text{Ш}(t) = s(t)\text{Ш}(t) = s(t)^{(0)} \tag{2.60}$$

i. e. the input signal. We have thus proven that, similarly to l_2-reconstruction (Sec. 2.4.2), IST (Sec. 2.4.4) and forward MaxEnt (Sec. 2.4.6), CLEAN does not modify measured data points, and its effect is non-trivial signal interpolation or extrapolation to unmeasured ones. However, a unique feature of CLEAN is that only a table of intensities at selected frequency points, $S^i_{\max}$, have to be stored to reconstruct spectra. It is of particular relevance for high-dimensional NMR, and is also extensively exploited in SSAlgorithm proposed in this thesis (Sec. 3.2.4).

The role of γ parameter is worth commenting. The subtraction of a ratio $\gamma \leq 1$ of full intensity at $\omega_{\max}$ reflects limited probability that subtracted peak is a real feature. In this interpretation, γ could even be made variable, depending on separation of the considered peak and other spectral peaks (noise): the more ambiguous spectrum, the safer (smaller) progress of the process. In practice, however, a trade-off with performance is made, and values $\gamma \approx 0.25 - 0.5$ are used. Larger values are known to cause artificial line splittings (Keeler 1984; Davies et al 1988). Small values are advised for spectra with broad, or overlapped lines, or very noisy.

The numerical efficiency of CLEAN may be disputed, since convergence, being relatively fast at the beginning, dramatically decreases at later iterations. Ignoring the fact that every iteration is fast and involves no Fourier transforms, point-by-point processing seems quite inefficient. A remedy suggested by (Högbom 1974) was to store all spectral points where deflections from mean intensity are significant, and update spectrum only at those points. Full search is only performed once all previously marked points are exhausted. SSAlgorithm proposed here (Chap. 3) exploits this idea even further to boost overall performance.

Similarly to MaxEnt (Sec. 2.4.6), first applications of CLEAN to NMR were aimed at elimination of truncation artefacts (Keeler 1984), resolution enhancement by deconvolution of instrumental lineshape (Davies et al 1988) and removal of phase-distortions in two-dimensional NMR (Shaka et al 1984). Shortly after sparse exponentially sampling was proposed by Barna et al (1987), CLEAN was also applied to NUS data[4], showing that it is more reliable way of enhancing resolution than extrapolation of truncated data (Barna et al 1988).

Barna et al (1988) showed that in the case of noisy data, the final result of CLEAN depends greatly on termination criterion, and advocated for use of MaxEnt in those cases. On the other hand, (Davies et al 1988) emphasised that both methods require careful choice of parameters to yield meaningful results for demanding data, and inclined towards use of CLEAN for the following reasons: (*i*) significantly lower computational and storage requirements, (*ii*) more intuitive parameters, and (*iii*) conceptual simplicity. Noteworthy, a fairly good solution to issue of termination point was proposed by Coggins and Zhou (2008), who suggested to employ additional noise stabilization criterion. The condition was quantified as follows:

$$\bar{\sigma}_j \leq (1+\tau)\bar{\sigma}_i \text{ for } i - 25 \leq j < i \tag{2.61}$$

where $\bar{\sigma}_j$ denotes noise level estimated in the jth iteration, and $\tau \approx 0.05$ is a stabilization factor. It should be emphasised that pionieering contribution of Coggins and Zhou (2008) was the first truly four-dimensional processing of high-resolution NUS data. However, their implementation follows closely the original algorithm, bearing the same deficiencies and moderate efficiency. One should also comment that Coggins' program name „FFT-CLEAN" is quite misleading, since FFT algorithm can be employed in virtually all reconstruction methods, and does not always provide expected performance gains for NUS data (see Sec. 3.2.3 and (Stanek and Koźmiński 2010)). Nowadays CLEAN seems one of very few algorithms capable of processing demanding high-resolution data. Surprisingly, this resembles similar heavy limitations encountered in 1980's for processing 2D spectra. In addition, it reproduces peak amplitudes with satisfactory linearity (Werner-Allen et al 2010).

It should be mentioned that CLEAN is also a more general idea of dividing signal into small and large components, as utilized in *distillation* process for forward MaxEnt method (Hyberts et al 2009). Its generality was also exploited in processing of non-gridded data,

[4]NUS and reconstruction methods were of limited use since it was not possible to program spectrometer hardware to sample non-linearly at that time (Barna et al 1988). Demonstrated were only potential time-savings since processed data were a subset of complete experiments, and not truly NUS.

including radially sampled in conjunction with back-projection reconstruction (Kupče and Freeman 2005), and randomly sampled without spatial restrictions (Kazimierczuk et al 2007). The relationship of SSAlgorithm to the semi-automatic model-based procedure of Kazimierczuk et al (2007) is discussed in more detail in Sec. 3.1.4.

One should also emphasise limitations of relatively unsophisticated scheme of CLEAN. Firstly, its incapable of processing spectra where no peaks are initially evident due to accumulation of NUS artefacts. Secondly, low sensitivity and/or intense NUS artefacts hinder the artefact suppression since greatly incorrect peak amplitudes are used for subtraction of „*dirty mask*" (Barna et al 1988). Moreover, these errors accumulate in subsequent iterations, and there is no mechanism to correct it. Finally, extraction of strong peaks is incomplete due to non-zero detection threshold (eq. 2.54), affecting denoising performance. SSAlgorithm proposed in this thesis addresses all these issues (see Sec. 3.3). Recently, some of these improvements were incorporated in the algorithm SCRUB (Coggins et al 2012).

2.4.4. Compressed sensing

Compressed sensing deals with the broad class of problems for which the solution is known to be *sparse*. The following regularizing term is used:

$$\mathcal{R}(S) = ||S||_0 \tag{2.62}$$

where $||S||_0$ is the number of non-zero components, often (erroneously) termed "$\mathcal{L}_0$-norm". The objective function is non-convex, thus multiple local minima may exist. In general, finding the sparsest solution is a combinatorial, *non-polynomially* (NP) hard problem. As a result, the approach seems intractable for all but very small problem sizes.

Recently, Candès et al (2006) and Donoho (2006) proved that for sufficiently sparse signals at a given level of underdeterminicity, there exists equivalence between finding the sparsest solution and l_1-norm minimization ($\mathcal{L}_1(S) = \sum_{k=1}^{M} |S_k|$). More formally, if the number of measurements K satisfies

$$K \geq C\mu^2 S \log M \tag{2.63}$$

where μ is the so-called incoherence between orthonormal bases in which the signal is expressed[5], M is the full problem size, S – the number of non-zero components in the sparse representation of the signal, and C is some positive constant, then the constrained $l_1$minimization yields *exact* sparse solution with overwhelming probability.

The theorem has profound consequences: 1) the minimized functional is convex. Therefore the solution is unique, and can be found in the polynomial time employing general-purpose or specialized optimization algorithms (linear programming). 2) *perfect* reconstruction of sparse signal can be found, i.e. *without any loss of information.* Allowed undersampling level depends on the sparsity S, and logarithmically on the number of features in the spectrum M. 3) no prior knowledge about the number, location, and amplitudes of nonzero components is assumed. Effectively, CS theory paved the way to replace post-sampling compression (such as in JPEG or MP3 formats):

full (dense) acquisition $\longrightarrow$ nearly sparse representation $\longrightarrow$ (lossy) compressed data set

with a novel scheme of

non-adaptive sparse acquisition $\longrightarrow$ l_1 min. $\longrightarrow$ decompressed sparse representation

[5]The signal is assumed to have a sparse representation in at least one of orthonormal bases. E.g., incoherence between Dirac comb in the time domain and Fourier basis has the smallest possible value of $\mu = 1$.

The signal is acquired in the compressed form, hence the name "Compressed Sensing".

The theory of CS is even more powerful, stating that l_1 minimization yields approximate solutions for the signals that does not have strictly sparse representations. Under certain assumptions regarding rank-deficient matrix of linear equations (e.g., eq. 2.27), the minimum of $l_1$functional, x*, obeys:

$$||x^* - x||_2 \leq C_0||x - x_S||_1/\sqrt{S} + C_1\epsilon \tag{2.64}$$

where x is the perfect solution (not necessarily sparse), x_S is its approximation preserving only S largest components, ϵ indicates error bounds of measured data, and C_0 and C_1 are some constants. In other words, this theorem puts indeterministic upper bounds on the quality of l_1-minimum (left-hand side), which mostly depend on sparsity of the signal (the first term) and measurement errors (the second term on the right-hand side). Quite striking feature of CS is that it extracts S largest components of non-sparse signal without prior knowledge of their locations. This powerful theorem makes assumptions on transformation matrix, which inherently depends on sampling strategy. It can be proved that randomized sampling pose least stringent requirements on minimal number of measurements, e.g.,

$$K \geq C\, S \log(M/S), \tag{2.65}$$

in order to hold required *restricted isometry property*. In this sense, randomized sampling represents near-optimal compressed sensing strategy of general, noisy signals that are believed to have sparse representation (Candès and Wakin 2008).

The advent of CS theory has triggered applications also in the field of processing of NMR signals. Particularly promising are assumptions of CS, which seem well-suited to multidimensional NMR spectra. The first application was shown by Drori (2007) to reconstruction of undersampled 2D HSQC spectra. However, wavelet transform was unnecessarily employed, affecting the clarity of the approach. Nevertheless, Drori demonstrated that *iterative soft thresholding* (IST) is an effective scheme for finding approximation to the sparsest solution. Although IST is a heuristic algorithm, it outperforms general solvers of $l_1$minimum for practical problem sizes. Simultaneously, Stern et al (2007) formally showed that the fixed-point of IST is also a minimum l_1-norm reconstruction. This important result ensures validity of IST reconstruction by connecting it with CS theory. Also, it suggests the way a stable deconvolution may be incorporated into IST scheme. Stern et al (2007) also emphasised the formal similarity of CS and MaxEnt, and compared their effiacy on the truncated signals. The slow convergence of IST was also demonstrated, and more sophisticated minimization algorithms, such as based on conjugate-gradient methods, were suggested to avoid premature termination of iterations, before the correct minimum is obtained. The comparison of numerical performance between $l_1$minimization and MaxEnt showed approx. 6-fold advantage of the latter; this result was attributed primarily to the advanced dedicated minimizer used in the Cambridge algorithm, and *not* to the intrinsic properties of l_1 functional.

Noteworthy, l_1 regularizing function was soon employed as an alternative to entropy by Hyberts et al (2009), however, without a clear connection to CS theory. Also, CS was employed to process non-uniformly sampled *direct* dimension data in 2D ultra-fast experiments performed using irregular gradient patterns (Shrot and Frydman 2011). The latter study facilitated the equivalence of l_1-minimization and simple IST processing, as well as the the relatively large sparsity of t_2 signals in spatially resolved spectra.

In the subsequent work of Holland et al (2011), l_1 minimization was applied for the first time to protein 3D spectra (HNCA and HN(CO)CA), and compared favourably to MaxEnt reconstructions at sparsity level of 20%. It was demonstrated that CS is capable of

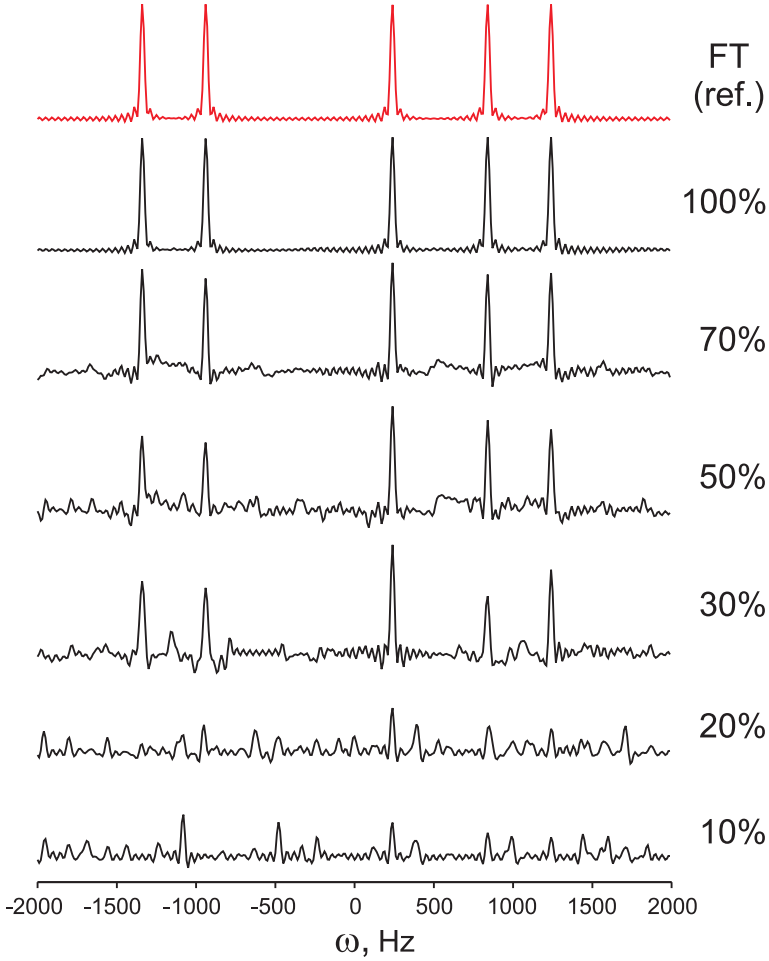


Fig. 2.11 l_1 reconstructions of nearly sparse spectrum containing five components of *equal* intensity and decay rate. *Top* panel shows the reference fully sampled FT spectrum, while other panels represent reconstructed spectra with decreasing (from top to bottom) sampling density. At some critical point the quality of l_1-reconstruction breaks down, showing meaningless results.

reconstruction of indistinguishable from fully sampled spectra, and performs well in terms of precision of extracted chem. shifts for both high and low S/N regimes. Simultaneously, Kazimierczuk and Orekhov (2011) suggested to employ *iterative re-weighted least squares algorithm* (IRLS) (Candès et al 2008) instead of IST (or l_1 minimization) for CS reconstruction, and demonstrated results on 2D HSQC and a far more complex 2D NOESY spectra. Contrary to l_1 minimization, IRLS allows to minimize norms < 1 (denoted $l_{p\to 0}$), and exhibits faster convergence than IST for complex signals. Despite $l_{p\to 0}$ norm is non-convex and may have more than one local minimum, the CS theory predicts that a smaller number of measurements is needed to obtain exact reconstruction (Chartrand and Staneva 2008; Chartrand and Yin 2008). Kazimierczuk and Orekhov also proved that CS correctly reproduces relative intensities, and arbitrary lineshapes shapes, including antiphase ones such as in DQF COSY. The issue of non-convex functionals was further investigated by Kazimierczuk and Orekhov (2012), and a detailed comparison between IRLS, IST, MDD and l_2-reconstruction was made on both synthetic and real NMR data. For most cases, IRLS yielded "perfect" reconstruction (with deviations from fully sampled reference below an arbitraty threshold) for the smallest input data set. It was also noted that NMR spectra containing noise and slowly decaying Lorentzian lines are not strictly sparse. This dramatically decreases the convergence of IST, and may lead to misinterpretation of resulting spectra. Fig. 2.11 shows l_1 reconstructions of a noiseless signal containing 5 truncated, exponentially decaying components, which has a nearly sparse FT spectrum (*top*). It is evident that the quality of reconstruction gradually degrades if insufficient data is provided in comparison to signal sparsity.

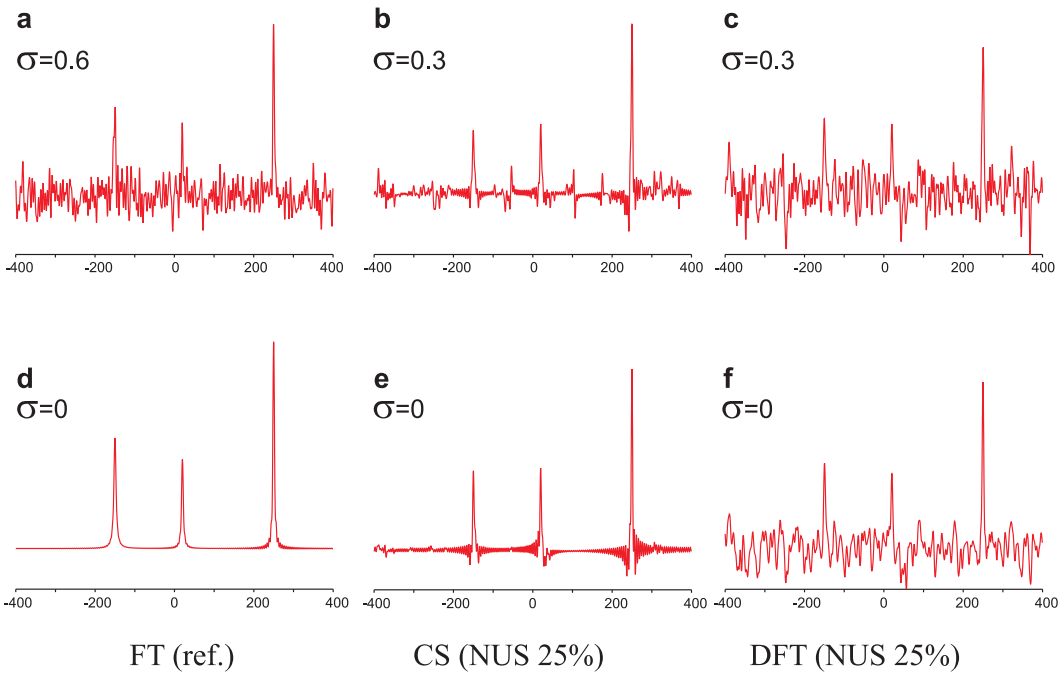


Fig. 2.12 Non-linear noise transformation in l_1-reconstructed synthetic spectra. Compared are fully sampled reference FT spectra (*a, d*), l_1- (*b, e*), and l_2-norm reconstructions (*c, f*) of 75% undersampled data. While l_1-reconstructions perform superior for noiseless signal (*d-f*), substantial amount of Gaussian noise causes difficulties in the proper recognition of genuine peaks (*b*). The noise was increased 2-fold for the fully sampled data set to account for the difference in number of samples.

According to the CS theory, in the case of severely undersampled data, only strong peaks are correctly reproduced while the smaller ones are suppressed. This explains why MDD proved superior to IRLS for 3D NOESY applications (Kazimierczuk and Orekhov 2012). Additionally, the non-linear transformation of noise invalidates the notion of S/N and may cause difficulties in distinguishing genuine signals from the background (Hyberts et al 2013). The problem of amplification of strong noise peaks, and suppression of weak ones is illustrated in Fig. 2.12. The enforced sparsity of the spectra yields superior results for noiseless data, however, difficulties arise for signals at critical sensitivity, since noise peaks may be misinterpreted as *false positives* (Fig. 2.12b).

Kazimierczuk and Orekhov (2011) showed that computational complexity of IRLS grows with the third power of the number of signal samples, K (or sparsity), making it quite impractical for large data sets. On the other hand, processing time of IST it is independent of K. Therefore, IST seems a practical solution for processing high-resolution, and high-dimensional spectra, as showed by Hyberts et al (2012) in application to 4D NOESY. The detailed study on the quality of 3D NOESY spectra obtained with a conservative 40% NUS and l_1-reconstruction was recently presented (Bostock et al 2012).

2.4.5. SIFT

Sparsity of spectral domain can be utilized in a more controlled way than what is performed by l_1-minimization. It was proposed to employ the knowledge of "dark" regions to support spectral reconstruction by explicitly enforcing zero spectral intensity (Matsuki et al 2009). This is possible in the Gerchberg (1974) - Papoulis (1975) iterative algorithm, which scheme is shown in Fig. 2.13. As can be appreciated from the scheme, known time-domain (measure-

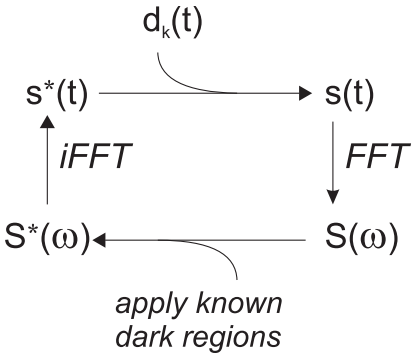


Fig. 2.13 The SIFT scheme. In the Gerchberg-Papoulis algorithm alternated forward and inverse FT are employed. Zero spectral frequencies and measured data points are reintroduced to spectrum and signal in every cycle.

ments $d(t)$) as well as frequency-domain information are incorporated in the reconstruction. The linearity of FT ensures that this information is treated equivalently. Effectively, a given fraction of missing measurements can be recovered by the same fraction of known spectral information. The name *"spectroscopy by integration of frequency and time-domain information"* (SIFT) clearly reflects the underlying idea.

It was argued by Matsuki et al (2009) that SIFT facilitates a very concrete and unambiguous information, avoids introduction of any specific model, and is, in general, a safer approach than IST. Since no threshold is imposed, there is no risk of suppressing small peaks such as in NOESY applications. However, Matsuki et al (2009) admitted that IST can generally provide better artefact suppression as sparsity is enforced more aggressively. It was also shown that SIFT converges rapidly, and is robust in terms of linear reproduction of signal intensities. Additionally, spectral quality degrades gently in the subcritical regime, i.e. if more measurements are missing than what can be recovered by dark spectral information.

It should be emphasised that SIFT makes quite practical and reasonable assumptions regarding NMR spectra. Often, dark regions can be plainly identified in spectra of lower dimensionality, e.g. 2D ^{13}C-HSQC, and transferred to their high-dimensional extensions. Also, in MAS solid-state NMR experiments, indirect time increment is rotor-synchronized to fold the spinning sidebands exactly onto the corresponding ceterbands. As a result, bandwidths exceed the extent of signal dispersion, leaving vast spectral regions bare. The advantages of SIFT were demonstrated in NUS 3D NCOCX spectrum featuring critical sensitivity (Matsuki et al 2010). Interestingly, the employed "dark" spectral information leads to increased S/N similarly as the noiseless data were measured, thus outperforming methods relying entirely on time-domain data (e.g. MDD).

In conclusion, it is worth emphasising the similarity between assumptions of SIFT and CS. While the latter facilitates most of spectral sparsity for enhancing reconstruction, the former follows a more conservative route. CS is definitely a more general and practical approach, however, may promote strong peaks at the expense of smaller ones if undersampled. SIFT is certainly safer in this respect.

2.4.6. Maximum entropy reconstruction

Images (spectra etc.) can be assigned a so-called configurational entropy, introduced to the information theory by Shannon (1948). A central assumption is that luminance quanta have a priori equal chance of being in any pixel. Then the number of ways of getting a particular configuration (image) can be approximated by entropy

$$S(\mathbf{u}) = -\sum_{i=0}^{N-1} u_i \log(u_i/U) \tag{2.66}$$

where $U = \sum u_i$ is the total number of quanta in the image (Press et al 2007, §19.7). The equivalent, compact form of

$$S(\mathbf{u}) = -\sum_{i=0}^{N-1} p_i \log p_i, \tag{2.67}$$

where $p_i = u_i/U$ can be interpreted as probability of the presence of luminance quantum in the given ith piece of the image, was originally proposed by Frieden (1972). The maximum entropy image restoration was first applied in radioastronomy (Gull and Daniell 1978), where the assumption about flat probability distribution of signals seems justified. It was argued that unconstrained maximisation of entropy leads to equal values of all pixels ($= e^{-1}$), and the only features that appear in the reconstructed maps are due to data constraints. In other words, MaxEnt only shows peaks if there is strong evidence in the measurements. Otherwise, MaxEnt yields uniform, featureless images (spectra). Also, resulting maps are positive and smooth if data errors are not underestimated.

Shore and Johnson (1980) proved axiomatically that MaxEnt method is the only consistent way of selecting a single image from many images which fit the measured data (see also Skilling (1984) for the accessible explanation). Thus, MaxEnt is maximally non-comittant with respect to the data, and does not introduce any correlations, beyond those required by data consistency. Other arguments in favour of the form of eq. 2.67 were given by Skilling (1984). Alternatively, instead of interpreting MaxEnt reconstructions as having the least amount of information still consistent with a measurement, one can simply treat entropy functional as a regularizing functional that ensures smooth and stable reconstruction from noisy data (Stern et al 2002).

Skilling and Bryan (1984) proposed to modify the entropy functional to a more useful (practical) form

$$S(\mathbf{U}) = -\sum u_i \left(\log(u_i/A) - 1\right) \tag{2.68}$$

where A (also termed *def*) is a used defined scale parameter representing the default (background) value of pixel intensity. The advantage of this form are simple analytical derivatives. The difficulty that arose in early applications of MaxEnt to NMR spectra (Sibisi et al 1984; Laue et al 1985b; Hore 1985; Hoch 1985) was that they are complex-valued, and even the real part is not necessarily positive. A number of workarounds have been used (Laue et al 1985a, 1986), however, the most elegant solution was proposed by Hoch et al (1990) and Daniell and Hore (1989). These two equivalent approaches, one extending the Shannon information formula to complex-valued spectra with differentiability conditions, and the second considering the entropy of an ensamble of spin$-\frac{1}{2}$ particles, lead to the entropy formula

$$S(\mathbf{u}) = -\sum_{i=0}^{N-1} \frac{|u_i|}{A} \log\left(\frac{|u_i|/A + \sqrt{4 + |u_i|^2/A^2}}{2}\right) - \sqrt{4 + |u_i|^2/A^2} \tag{2.69}$$

Maximum entropy functional is highly non-linear, and necessitates the use of iterative methods to find the solution. Notably, both entropy and χ^2 constraint term are everywhere convex, thus the unique minimum of the functional exists. Unfortunately, due to pronounced nonlinearity entropy functional fails to minimize using general-purpose line-search methods such as conjugate gradients or variable metric, which basically assume the quadratic form local approximation (Press et al 2007, §19.7.2). Skilling and Bryan (1984) proposed a sophisticated, but successful, algorithmic scheme of choosing the best search directions in appropriately chosen 3-dimensional subspaces. The later modification of the "Cambridge" algorithm proposed by Schmieder et al (1993) is now most commonly used, and is implemented in Rowland NMR Toolkit (RNMRTk, http://rnmrtk.uchc.edu/). A quite surprising fact, yet missing

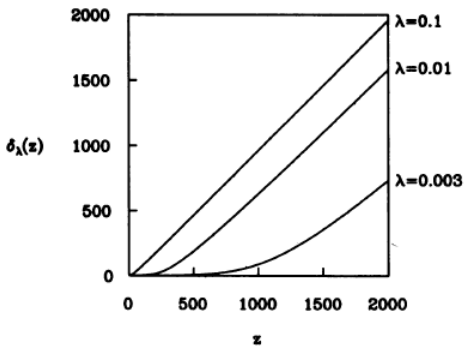


Fig. 2.14 Transfer function δ_λ that determines the features of MaxEnt reconstruction, at least in certain cases. Increased significance of entropy term (small λ) results in larger nonlinearity. Reprinted from Donoho et al (1990).

a clear explanation, is that "Cambridge" algorithm requires 30–50 iterations, irrespectively of the complexity of input data (Hoch and Stern 1996; Press et al 2007). An interesting aspect of the algorithm is that it exploits analytical derivatives of both entropy and χ^2 constraint to adjust λ during minimization. In effect, the desired correct constraint value is obtained in a single run. The highly ineffective iterative scheme of finding λ that corresponds to the desired data consistency is thus avoided (Skilling and Bryan 1984).

In general MaxEnt reconstructions must be found numerically. However, there is a special case of fully sampled data, $K = N$, where analytical solution can be derived (Donoho et al 1990; Hoch et al 1990). Although not strictly applicable to other cases, it gives insights into how the method works, and how it is affected by parameters A (*def*) and C_0 (*aim*). Parseval's theorem allows to replace χ^2 statistic by the sum $\sum_{i=0}^{N} |\hat{S}_i - S_i|^2$ where $\hat{S}_i$ and S_i are the MaxEnt and DFT spectra, respectively. At the critical point of objective function

$$0 = \nabla Q = \nabla S - 2\lambda(\hat{S} - S) \tag{2.70}$$

Let us define the following function

$$\delta_\lambda(z) = z + \frac{1}{2\lambda}\frac{\partial S}{\partial z} \tag{2.71}$$

By definition, $\delta_\lambda(S)$ fulfills eq. 2.70 and is thus the desired MaxEnt reconstruction. The transfer function δ_λ predicts how the DFT spectrum of measured data is transformed, in a point-to-point fashion, into a MaxEnt spectrum. As can be inferred from Fig. 2.14, MaxEnt suppresses weak spectral points more than intense ones (phase is not affected). Nonlinearity is an inherent feature of MaxEnt, and most pronounced for small values of λ, i.e. low significance of constraint term. Linear reconstructions can be obtained by increasing λ, however, at the expense of artefact suppression of DFT spectrum. In contrast, underestimated λ (or, equivalently, overestimated noise C_0) leads to flat featureless spectrum since data are largely ignored in the objective function. In general, λ is adjusted to obtain the reasonable value of χ^2; the latter is estimated based on noise in the input data (Mobli et al 2007). The nonlinear effects are clearly present on the synthetic spectra in Fig. 2.15, where noise is more suppressed close to the baseline than when superimposed onto peak. The second parameter that influences MaxEnt reconstruction is *def*, or A – the scale parameter present in the entropy functional (Eq. 2.69). One can notice that in the large *def* limit

$$\lim_{||\mathbf{u}||_2 A \to 0} S(\mathbf{u}) = -\text{const} \times \frac{1}{A^2} \sum_{i=0}^{N-1} |u_i|^2 = -||\mathbf{u}||_2 / A^2 \tag{2.72}$$

i.e. entropy functional becomes proportional to the spectrum power divided by def^2. As explained in Sec. 2.4.2, such a functional leads to DFT of zero-filled data. On the contrary,

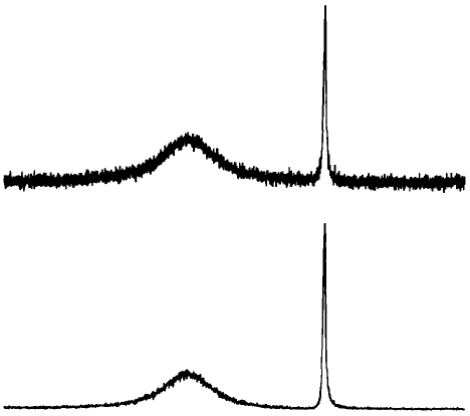


Fig. 2.15 DFT spectrum of two synthetic signals of considerably different linewidths (*top* panel), and the corresponding MaxEnt reconstruction (*bottom* panel). While Gaussian noise is uniformly distributed over entire DFT spectrum, MaxEnt scales down smaller intensities, leaving increased noise on the peak. Reprinted from Hoch and Stern (1996).

too small values lead to spiky baselines with amplified noise peaks. In general, *def* determines the level at which nonlinearity of reconstruction becomes significant, however, its role is less pronounced than *aim*. The "practical" rule is to use the value larger than noise but smaller than the smallest relevant spectral features (Hoch and Stern 2001).

MaxEnt processing of multidimensional spectra encounters a specific difficulty that, given the constant *aim* for constraint term, C_0, the consecutive rows (planes, etc.) in the indirect dimension(s) will undergo a transformation with variable λ. This means that depending on the data, the balance between entropy and constraint term varies, and may lead to distorted lineshapes. The simplest remedy is so-called *holistic* reconstruction, however, because of enormous storage requirements it seems quite impractical. The alternative solution, known as *constant-*λ algorithm, was proposed by Schmieder et al (1997). In this approach, a representative row is first transformed with a given *aim*. Next, a resulting value of λ is kept constant for all rows (planes, etc.) in a multidimensional spectrum. As an effect, the degree of nonlinearity is uniform across the entire spectrum.

Thanks to the advanced optimizer, MaxEnt reconstruction typically converges within $\sim$40 iterations, each involving 6–8 DFTs and a number of inner products. Additionally, efficient computation requires storage of 16$\times$ final spectrum size. While these requirements seem moderate for two- and three-dimensional spectra, four-dimensional processing poses a true challenge (Mobli et al 2010), despite availability of parallel distributed MaxEnt implementation (Li et al 1998).

MaxEnt represents a versatile method for spectral analysis. Initially, it was employed for resolution enhancement. Truncated data records are reliably processed and the method clearly outperforms linear prediction (Stern et al 2002). The more relevant application is currently NUS, where MaxEnt is usually employed to accelerate acquisition of spectra for resonance assignment (Rovnyak et al 2004a; Frueh et al 2006b). NOESY applications are possible, but necessitate somewhat involving *in-situ* calibration of intensities to compensate nonlinearity (Schmieder et al 1997).

One interesting feature of MaxEnt is its capability of stable deconvolution from a known function. Deconvolution of signal decay (Sibisi 1983) seems impractical as variations of decay rates are usually significant. In contrast, deconvolution of well-defined known scalar couplings can lead to simplification of spectra and true sensitivity enhancement. Examples include deconvolution of $C^\alpha - C^\beta$ (Shimba et al 2003), $C^\alpha - C'$ in carbon-detected experiments, or C–C couplings involving ^{13}C in methyl groups (Hoch and Stern 2001; Jordan et al 2006). Noteworthy, *virtual deconvolution* can be incorporated in other methods employing *inverse theory* approach, i.e. l_1- and l_2-norm minimization.

Recently, a variation of MaxEnt, referred to as *forward Maximum Entropy* (FM) was

proposed (Hyberts et al 2007). In this approach exact match between the *inverse* DFT of reconstructed spectrum, and measured data is enforced. Therefore, constraint term and *aim* are no longer necessary. In essence, the entropy is used to interpolate signal at points omitted in NUS schedule. Original entropy functional ($S(\mathbf{f}) = -\sum f_i \log f_i$), without *def* parameter, is used. FM was demonstrated to yield linear reproduction of signal amplitudes, and applied to demanding spectral analysis of metabolite mixtures (Hyberts et al 2007) and protein 3D and 4D NOESY spectra (Hyberts et al 2009). A significant drawback of FM are increased computational demands which necessitate the use of high performance CPU clusters.

2.4.7. Multidimensional decomposition (MDD)

Multidimensional decomposition (MDD) (Carroll and Chang 1970; Harshmann 1970), a. k. a. *multi-way decomposition*, *canonical decomposion* (Caroll and Pruzansky 1984) or **para***llel* **fac***tor analysis* (PARAFAC) (Bro 1997), is a quite different approach to analysis of NMR data than presented previously. In contrast to very general methods such as l_1, l_2-minimization or MaxEnt, MDD employs a model of data, in which the multidimensional spectrum (or time-domain signal) is decomposed into a small set of vectors. The spectrum S is assumed to be a sum of R outer products of these vectors, e.g. in a 3D case:

$$\boldsymbol{S} = \sum_{r=1}^{R} a_r \, \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r + \boldsymbol{E} \tag{2.73}$$

The normalized vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are commonly called *components*, *modes*, *loads*, or *shapes*, however, they do not necessarily describe any peak shape. They can represent both time- and Fourier domain data, or non-frequency dimensions such as exponential decays (Korzhnev et al 2001), J-modulation, or a time-dependent signal amplitude upon chemical reaction or temperature variation (Mayzel et al 2014). Mixed models, with some *shapes* in time and others in frequency domain, are allowed as well (Orekhov et al 2001). Moreover, *shapes* may contain multiple peaks, and may even represent merged two-dimensional spectral planes (Tugarinov et al 2005b). As a result, MDD model is versatile enough to describe both typical data, or series of experiments such as relaxation measurements or time-dependent spectra, for which FT is not applicable. Also, as described below, a single *hyperdimensional* MDD model may be employed to a set of different experiments, assuming that they share common dimensions.

It should be emphasised that, contrary to Bayesian (Bretthorst 1990) or *maximum likelihood methods* (Chylla and Markley 1993, 1995), MDD model is fairly non-restrictive. *Shapes* may represent arbitrary lineshapes, including unresolved J-multiplets, overlapped or non-Lorentzian shapes.

It is argued that the residual term $\boldsymbol{E}$ should contain virtually all noise, since (axiomatically) uncorrelated noise does not fit to the model (Orekhov et al 2001). Nevertheless, Maciejewski et al (2009) raised criticism that the issue of influence of noise on the stability of MDD remains somewhat under-investigated.

Certainly, MDD model suites well the majority of NMR experiments, since lineshapes in different dimensions are almost always independent (Luan et al 2005). MDD is unique for dimensionality $D \geq 3$ (Kruskal 1977). In a 2D case, there exists rotational ambiguity, which manifests in *mixing* of components (Orekhov et al 2001). Quite importantly, this phenomenon also occurs in 3DD if one vector is degenerate for two or more components. As it hinders the analysis of shapes, some automated solutions have been proposed (Gutmanas et al 2002). Noteworthy, *shape mixing* does not affect the spectral reconstruction from all *shapes*, which is the primary aim for NUS data.

The original application of 3DD was aimed at resolving overlapped peaks in uniformly sampled ^{15}N-edited NOESY by utilizing the fact that H^{NOE} vectors are distinct for (H, H^N) pairs (Orekhov et al 2001). Also, dramatic compression of spectral data and simplified analysis was emphasised. The possibility of NUS was also recognized and soon exploited for resolution enhancement (Orekhov et al 2003). In the case of *sparse* data the MDD model was slightly modified, leading the following penalty function (Ibraghimov 2002):

$$Q = |\boldsymbol{G} \circ (\boldsymbol{S} - \sum_{r=1}^{R} a_r \, \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r)|_2 + \lambda \sum_{r=1}^{R} (a_r)^2 \tag{2.74}$$

Here, the *sparsity* matrix has elements $G_{ijk} = 1$ if a data point was measured, and 0 otherwise, $\circ$ stands for Hadamard (element-wise) matrix multiplication, and $|\cdot|_2$ denotes sum of squares of elements of a matrix. The last term is to suppress large differences in the amplitude of components and boost convergence (Ibraghimov 2002). Its role is regulated by Tikhonov regularization factor λ (Tikhonov and Samarskij 1990).

If the *shapes* are of length I, J, K, the model contains $R\,(I + J + K - 2)$ variables. Thus, the decomposition is well determined if

$$\theta \cdot I \cdot J \cdot K \geq R\,(I + J + K - 2) \tag{2.75}$$

where θ stands for sparsity of input data. For typical 3D NMR spectra it should not be lower than $10 - 20\%$, which was determined by monitoring l_2 norm of difference of reference full data and MDD reconstruction from its subset (Orekhov et al 2003). It was argued that for NOESY experiments the limiting factor is rather sensitivity. However, minimum sparsity is an *implicit* limitation for resolution enhancement in other applications.

A crucial parameter for MDD is the number of components, R. It should be not smaller than expected number of *shapes*, and slightly overestimated (by 5-20%) values are recommended (Orekhov et al 2001). Usually, it is relatively straightforward to estimate R, e.g. by counting amide (H,H^N) resonances in ^{15}N-HSQC (for ^{15}N-edited NOESY or H^N-detected sequential assignment experiments), or in assigned ^{13}C-HSQC (for ^{13}C- or ^{13}C,^{13}C-edited NOESY). Noteworthy, in 3D NOESYs diagonal and cross-peaks are represented by a single *shape*, greatly simplifying the prediction of R. In contrast, in 4D NOESYs this is quite troublesome, since cross- and diagonal peaks share only two dimensions, and must be assigned separate *shapes*. In this case, the model is hard to establish, and usually converges to 4D spectrum with underestimated or missing cross-peaks (Luan et al 2005). For this reason, 3D processing of 4D NUS data is employed (Tugarinov et al 2005b).

Contrary to *sparsity* and number of *shapes*, the specific choice of Tikhonov regularization factor λ is not critical for reconstruction. The value of $\lambda = 0.005$ is recommended for most applications (Orekhov et al 2003).

As admitted by Orekhov et al (2003), the penalty function (eq. 2.74) is non-convex and the convergence to a global optimum cannot be formally assured. This belief is rather based on the long experience in application of MDD to various NMR data. It is also known that the convergence of the problems similar to that of Eq. 2.74 is intrinsically slow (Orekhov et al 2003), requiring at least 10^3 iterations. The algorithm was described in detail by Orekhov et al (2001), and involves some tricks to enhance convergence. It employs PARAFAC (Harshmann and Lundy 1984), an iterative alternated least-square algorithm to minimize penalty function of Eq. 2.74. The key point here is that only one vector is modified at a time, while all others are kept fixed. In the subsequent iterations, the vectors are permuted. Monotonous convergence is ensured since in every step the penalty function is minimized. Noteworthy is the initial

approximation for PARAFAC algorithm, proposed by Ibraghimov (1999) for the case of $R \leq \max(I, J, K)$. Also, *Tucker3* compression (Tucker 1966), which is valid if $R \leq \min(I, J, K)$, is employed in a modified PARAFAC algorithm to yield reasonable approximate solution of the original problem (Kroonenberg and Leeuw 1980). Vector orthonormality conditions and *singular value decomposition* are extensively exploited in this step. The dramatic speed-up of calculations follows the fact that the matrix compression often helps PARAFAC to converge in a few iterations. It should also be emphasised that MDD is amenable to parallelization: the data is divided into pieces in the directly detected dimension and processed independently. Ultimately, obtained *shapes* are used to reconstruct full time-domain data. They may be subsequently processed with FT using apodizing functions, zero-filling, etc. The scheme of MDD processing of 3D NUS data is shown in Fig. 2.16.

MDD yields linear reconstructions and tolerates high dynamic range (Luan et al 2005). Thus, it is well suited to NOESY applications (Orekhov et al 2003). The model is particularly beneficial for 3D NOESY spectra: the fact that cross-peaks and strong diagonal signals share *shapes* in ^{15}N and H$^{\mathrm{N}}$dimensions greatly supports the MDD reconstruction. The fidelity of the extracted structural information was verified (Gutmanas et al 2002). Particularly spectacular were applications to 4D ^{13}C,^{13}C-edited NOESY, which enabled structure determination of 82 kDa MSG protein (Tugarinov et al 2005a,b) or 283-residue membrane protein VDAC-1 (Hiller et al 2008a, 2009). However, 4D NOESY data sets must be processed with 3DD, which entails relatively high demands on sparsity ($\theta \geq 0.2 - 0.3$). As a result, only moderate

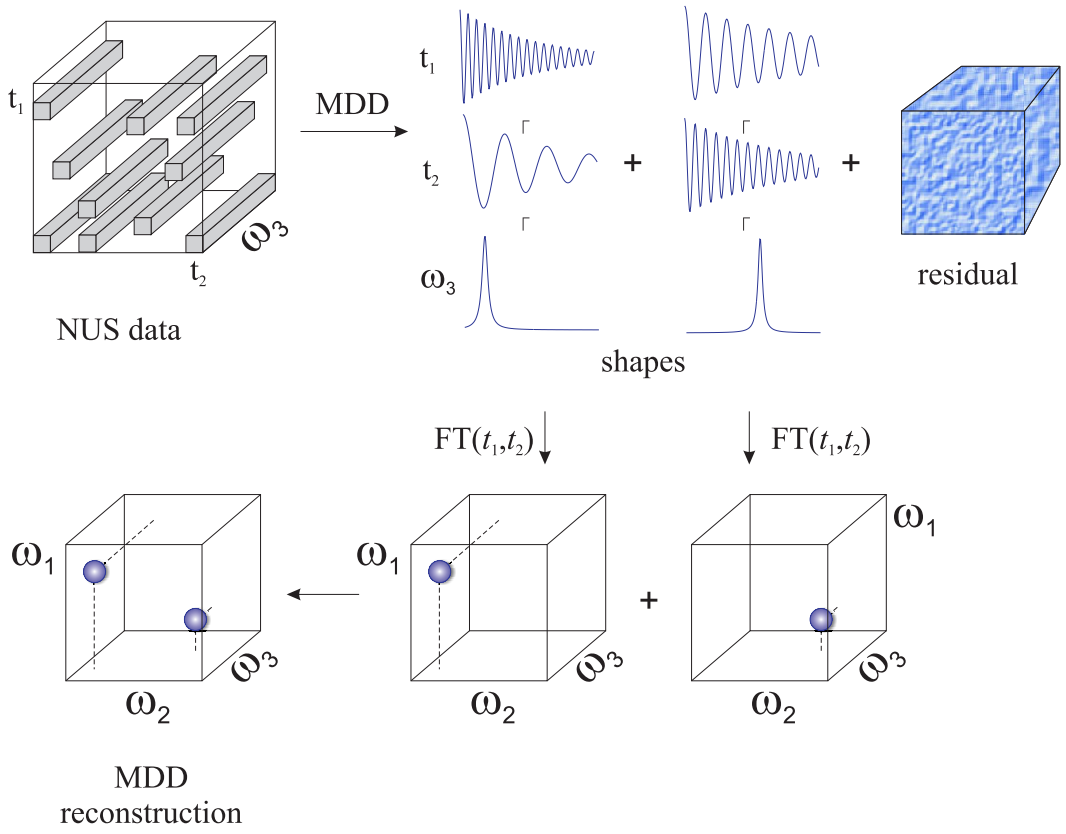


Fig. 2.16 The scheme of MDD processing of 3D NUS data. First, uniformly sampled t_3 interferograms are processed using FT (*not shown*). Subsequently, MDD generates a set of R (here two) components, each comprising three 1D vectors (for t_1, t_2 and ω_3). Fianlly, the shapes can be Fourier transformed to yield spectral reconstruction.

resolution enhancement is possible in indirect dimensions.

MDD generates relatively simple output, prone to automatic analysis and spectral assessment. This feature was exploited in a novel target-oriented NUS strategy, called *targeted acquisition* (TA) (Jaravine and Orekhov 2006). Prior to experiment specific targets must be defined, e.g. expected number of peaks or precision of chemical shifts. These parameters are monitored almost at real-time, allowing to stop the experiment once all targets are met. Effectively, sparseness level need not be assumed *a priori*, leading to optimal time allocation: More sensitive experiments and less demanding samples occupy less spectrometer time, which can be reallocated to less sensitive spectra.

MDD model can also be extended to describe a number of spectra sharing common dimensions. E.g., nine-dimensional model incorporates data from six 3D experiments: HNCO, iHNCA, iHN(CA)CB, HN(CO)CA, HN(CA)CO and CBCA(CO)NH. The approach referred to as *hyperdimensional* (HD) MDD (Jaravine et al 2008) allows to conveniently gather all information collected using different experiments. HD approach allows to establish remote correlations, provided that common dimensions are sufficiently resolved. E.g., a pair of HN(CO)CA and HN(CA)CO enables indirect correlation of C^{α}–C'. While this works satisfactorily for small folded proteins, difficulties arise for proteins exhibiting extensive overlap in ^{15}N-HSQC, e.g. IDPs (Jaravine et al 2008).

The real power of HD MDD is embedded in transferring information on *shapes* in shared dimensions from sensitive to insensitive experiments. This enables further time-savings since the entire HD MDD model converges at lower overall sparsity (Jaravine et al 2008). The HD approach is usually coupled with TA to monitor obtained spectral parameters in real-time. Also, automatic assignment routines are employed (Wong et al 2008) to verify completeness of sequential (or side chain) assignment. The latter reflects the ability to establish peak connectivities, not only raw number of peaks present in spectra, and is thus a correct measure of success of the protocol.

Noteworthy is also another example of co-processing shown by Hiller et al (2009). 4D ^{13}C,^{15}N-edited NOESY is an inherently diagonal-free spectrum, posing difficulty to converge MDD model. However, if (H^{N},^{15}N) shapes are transferred from simultaneously processed sensitive ^{15}N-HSQC, 4D spectrum can be reliably reconstructed from very sparse data.

It was also shown that in specific cases the MDD model can be modified to benefit from auto-regressive property of some NMR signals (Jaravine et al 2006). E.g., the *recursive* MDD (R-MDD) approach is applicable to HNCO, HN(CO)CA or HNCA. The incorporation of auto-regressive property allows to significantly reduce data sparsity. Since recursive MDD factorizes *shapes* into shorter vectors of length d, the effective number of unknowns reduces from K to $(d-1)\log_d(2K)$. In the cases where R-MDD is valid, resolution can be enhanced in logarithmic experimental time, instead of usual linear dependence.

In conclusion, MDD represents a very useful tool for automated analysis and optimal non-uniform sampling of NMR spectra. Although the introduction of a parameterized model might be considered a disadvantage, it also provides a convenient tool for utilizing prior knowledge, e.g. on dependencies between different spectra or expected number of peaks.

Chapter 3

Signal Separation Algorithm

3.1. Algorithm description

3.1.1. Overview

The premise for designing a new algorithm are well recognized deficiences of CLEAN (sec. 2.4.3). As its successor, the proposed algorithm is aimed at (*i*) optimizing NUS artefact suppression, (*ii*) improved handling of crowded and noisy spectra, and (*iii*) increased efficiency in processing and memory management, both required to deal with high-resolution and high-dimensionality spectra.

Fig. 3.1 shows the general scheme of *S*ignal *S*eparation *A*lgorithm (SSA). Provided that

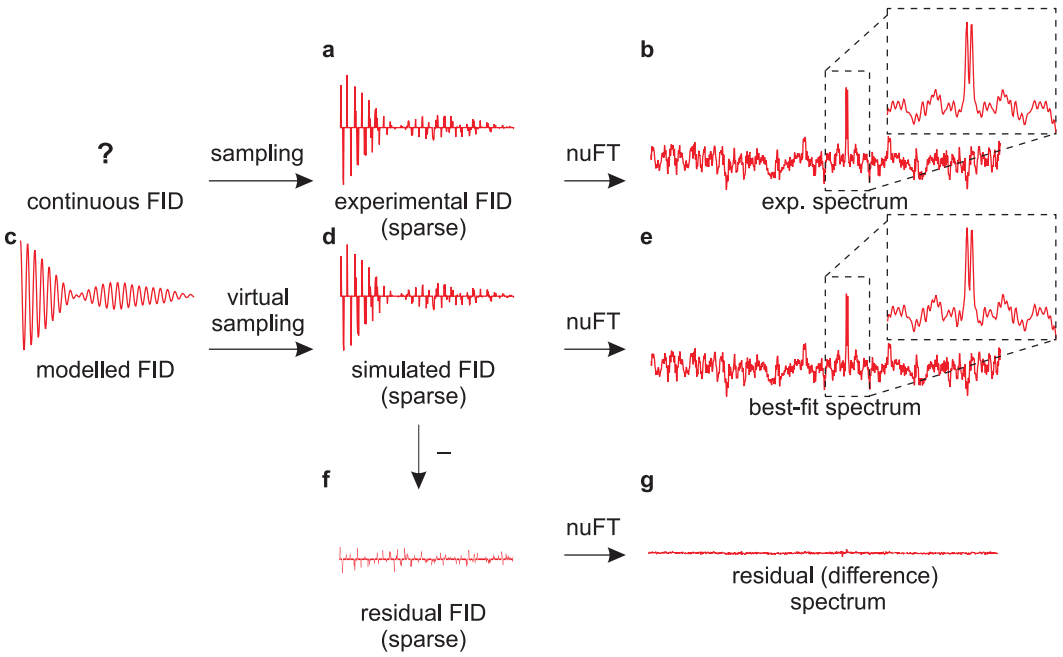


Fig. 3.1 Flowchart illustrating the principle of peak fitting in the frequency domain using a model for the time domain signal. Given a peak is localized in the z.-a. FT spectrum (*b*), the relative amplitudes of the frequency components in the model FID (*c*) are varied in order to obtain an identical line shape in the synthetic NUS z.-a. FT spectrum (*e*). The obtained signal model (*c*) contributes to "clean" spectrum, whereas the residual (difference) signal (*f*) gives rise to residual spectrum. Thus, the latter contains uncorrelated experimental noise as well as residual artefacts stemming from imperfect fitting.

the systematic portion of data has a nearly sparse spectral representation, SSA fits its time-domain model (panel c) so to minimize l_2-norm of the difference (panel f) between experimental (a) and synthetic signal (d) at the sampled points. Equivalently, the difference, a. k. a. *residual* spectrum (panel g) is supposed to contain no significant features, and only evenly distributed noise. It is worth emphasising that generalized FT is employed, and all known properties apply (Kazimierczuk et al 2012). For example, since z.-a. FT is a linear operation, fitting of *sparse* time-domain data has an intuitive corresponding representation in the Fourier domain (panels b and e).

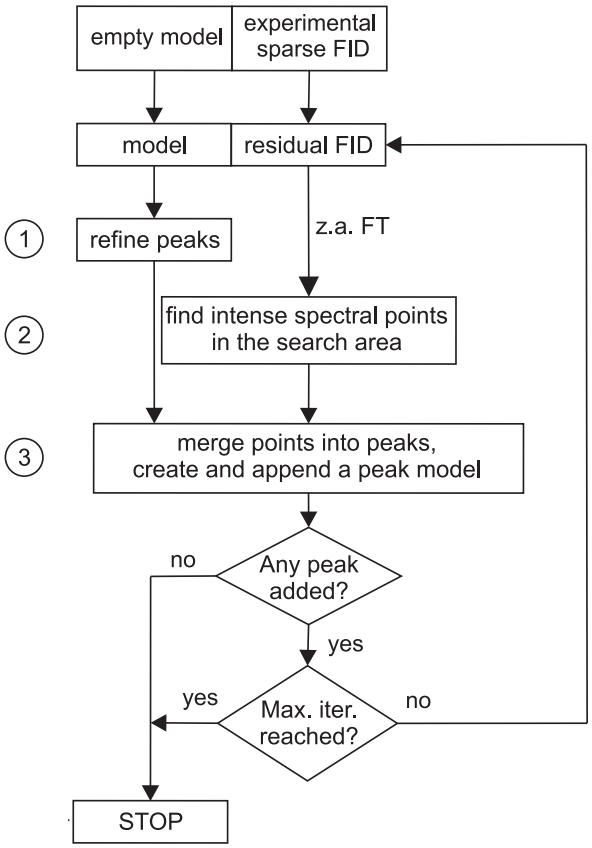


Fig. 3.2 The flowchart of SSAlgorithm. Initially, the empty model is assumed and experimental sparse data is used as residual FID. In the consecutive iterations, the cycle begins with refinement of peak models to resolve dependencies between them, and to rectify models established in the previous iteration (*step 1*). Subsequently, residual FID is Fourier transformed. Intense spectral features are searched in the predefined spectral areas (*step 2*). These spectral points give rise to newly created peaks, which are parametrized in the *step 3*. If there are any new peaks, the extended model and residual FID are used in the next cycle. Otherwise, the algorithm terminates.

The model is built-up gradually in the iterative scheme shown in Fig. 3.2. Peaks are recognized in the z.-a. FTresidual spectrum (*step 2*), then parametrized and appended to the model (*step 3*). Most likely, i.e. most intense peaks are processed first to minimize the risk of fitting to NUS artefacts. This approach implies a separate fitting of every resonance, which is beneficial for performance reasons. However, non-orthogonality of FT basis functions, associated with an incomplete sampling, causes signal interference in the spectral domain. All peaks are non-local to a certain extent, thus individual fits are affected by the presence of other peaks. To alleviate this issue a refinement procedure, repeated *before* searching for new peaks (*step 1*), is proposed. Suffice to say here, the peaks are effectively decoupled from each other, which justifies the algorithm's name.

The SSA model is general and explicitly sparse. It is created without user intervention and *a priori* assumptions about the number of signals, but using only objective criteria. In contrast to CLEAN, it uses only one parameter: relative detection threshold α. Also, as demonstrated in sec. 3.3, it is less prone to errors due to *speculative* thresholding, i.e. with an underestimated α. In the following sections, a more detailed description of the involved procedures is given.

3.1.2. Peak searching

Peaks are recognized on a variable thresholding basis, where the latter is associated with an estimated noise level. More specifically, a peak must contain at least one spectral point of intensity $I \geq \alpha\,\hat{\sigma}$, where a default $\alpha = 10$ is known to be sufficiently safe to avoid detection of artificial or noise peaks.

The noise estimate, $\hat{\sigma}$, is calculated using a *median* of all spectral points, which is evaluated independently for every $(N-1)D$-dimensional subspectrum for a directly detected point. To decrease associated overhead, 40 $(N-1)D$-dimensional subcubes, spanning $\frac{1}{10}$ of spectral window in every indirect dimension, are randomly selected from entire spectral subdomain.

Also note that, compared to a standard deviation, a median of amplitudes is less susceptible to the presence of intense outliers such as genuine peaks. Since only a tiny fraction of spectral volume is occupied by peaks, their influence on a noise estimate can be neglected. The relation of median and standard deviation can be derived exactly for Gaussian distribution. The cumulative distribution function (CDF) for an absolute value of a Gaussian random variable of zero mean is

$$\mathrm{CDF}(x) = \mathrm{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) \tag{3.1}$$

From a definition, the median m fulfills: $\mathrm{CDF}(m) = \frac{1}{2}$. Therefore

$$\frac{m}{\sqrt{2}\sigma} = \mathrm{erf}^{-1}\left(\frac{1}{2}\right) \tag{3.2}$$

$$\sigma = \frac{m}{\sqrt{2}\,\mathrm{erf}^{-1}(\frac{1}{2})} \approx 1.48\,m \tag{3.3}$$

i.e. a simple proportionality can be used to estimate a standard deviation of noise σ using a median of absolute values of spectral intensities m.

Increased threshold parameter $\alpha = 10$ accounts for the limited precision of noise estimation using a partial spectral domain, a median, and the deviations from uniform noise (and artefact) distribution. Theoretically derived Gaussian distribution of noise and artefacts is not assumed since multiple factors may contribute, e.g. a sampling schedule generation approach. Therefore, the used threshold is rather heuristic and has no immediate probability interpretation.

Notably, the lack of significant spectral features is used as a primary termination criterion. The noise stability criterion employed in the Coggins' implemenation CLEAN (Coggins and Zhou 2008) was disregarded as it generally leads to premature termination of this slowly converging algorithm. As noticed by Stern et al (2007), a similar issue arises also for IST. We thus prefer to perform an exhaustive peak search even if at the expense of computational performance.

3.1.3. Generic raster models

At the subsequent stage, the potential peaks are verified again vs. the current noise threshold since the residual spectrum is changed during this procedure. Next, peak areas are determined by joining adjacent spectral points which intensity exceeds an *expansion threshold*, $4\,\hat{\sigma}$. Note that created areas are fully-dimensional, i.e. including directly detected dimension. Peak areas are disjoint as points previously assigned to any other peak are discarded.

Once a peak area Ω is found, a generic model comprising all included frequency components is established: $g \equiv \sum_\Omega g_i\,\delta(\omega - \omega_i)$. This model is equivalent to the „replica" in

CLEAN, but derived in a quite different way. The cental assumption is that model g should, after inverse FT to the time-domain, sparsifying and discrete forward FT, faithully reproduce the peak shape in the residual spectrum (the so-called *aim*, a):

$$g(\omega) \xrightarrow{\text{iDFT}} g(t) \xrightarrow{\text{NUS}} \tilde{g}(t) \xrightarrow{\text{DFT}} r(\omega) \tag{3.4}$$

Noteworthy, real-valued model g can be used for recovery of complex-valued time-domain signal $g(t)$ if sufficiently fine digitized ($M \geq 2N$). This operation is equivalent to Hilbert transform (Bracewell 2000, §13).

The mismatch between the result $r(\omega)$ and expected shape $a(\omega)$ is used to modify the model:

$$g_i \leftarrow g_i \times \begin{cases} \frac{2}{1+(\frac{r}{a})^2} & \text{if} \quad \text{sgn}(r) = \text{sgn}(a) \\ \sqrt{2} & \text{if} \quad \text{sgn}(r) \neq \text{sgn}(a) \end{cases}$$

The above correction factor f is deliberately chosen to simultaneously satisfy 3 criteria:

1. $f = 1$ if $r_i = a_i$, i.e. the agreement has been obtained;
2. $f = 2$ if $r_i = 0$ (or any tiny numerical value),
3. $f \approx 1 - \frac{r}{a}$ if $|\frac{r}{a} - 1| \to 0$, i.e. an approximately linear correction is applied to the model g, depending on a ratio r_i/a_i, if the latter is close to 1.

Noteworthy, the correction quickly downscales the model if $|r_i/a_i| \gg 1$, and gently amplifies it otherwise. The stability is ensured by model area selection, i.e. where a_i has non-negligible values.

The procedure of model adjustment is iterative. Initially, $g = a$, and a consistent sign of model amplitudes g_i is enforced. Also, a rough scaling is applied first to account for intensity scaling embedded in the cycle of Eq. 3.4. Then adjustement cycles are performed until r.m.s. difference between a and r stabilizes within limits of 0.0025 (relative) between consecutive cycles, or unexpectedly increases. In practice, the model converges in approximately 100-200 cycles. Note that a single cycle involves only inverse and forward DFT from/to a small spectral region occupied by a peak.

Once a peak model is established, the corresponding time-domain representation (sparsified), $\tilde{g}(t)$, is subtracted from the residual FID. As a result, the artefact level supposedly decreases before next peaks are processed.

3.1.4. Fitting analytic models

In the cases when the signal in all dimensions can be reasonably approximated by mono-exponentially decaying sinusoids, a parametric model-fitting approach may prove powerful in retrieving of missing information. Previously, maximum-likelihood principle was employed to (*i*) extrapolate signals in constant-time experiments, (*ii*) resolve overlapped peaks in multi-dimensional spectra of limited spectra resolution, and (*iii*) reliably extract signal parameters, e.g. for relaxation studies (Chylla and Markley 1993, 1995; Chylla et al 1998).

Analytical signal models were also incorporated in the CLEAN-like algorithm for rough suppression of NUS artefacts (Kazimierczuk et al 2007). This approach has, however, a few of potential pitfalls:

1. peaks are fitted in the frequency domain with assumed Lorentzian shapes,
2. fitting procedure is manual,

3. no alternative was proposed if peaks overlap to the extent that they cannot be stably deconvolved.

Of a particular relevance is the first issue, since peak shapes in the Fourier domain are affected by convolution with the point response (including NUS and truncation effects), and often broadened by use of decaying sampling probability or non-Lorentzian apodization functions.

To avoid these issues we have proposed to fit sparse NUS data with mono-exponentially decaying *time-domain* signals

$$s(t;\ A,\omega,R) = A\exp(i\omega t - Rt). \tag{3.5}$$

In order to approximately decouple fits of particular peaks, the goodness-of-fit is, however, evaluated in the real-part of Fourier domain in the signal area Ω:

$$Q(A,\omega,R) = \sum_{\omega_m\in\Omega}\left(S_m - \mathrm{Re}\left\{\sum_{k=1}^{K} s(t_k;\ A,\omega,R)\exp(-i\omega_m t_k)\right\}\right)^2 \tag{3.6}$$

where S_m are elements of residual spectrum obtained by DFT over K signal samples. Effectively, all factors contributing to peak shapes are taken into account, since experimental and synthetic data are processed equally. Later, we have proposed to incorporate directly detected dimension into the analytic peak models, similarly as for MDD model (Stanek et al 2012). Two parameters are added to models $(\omega_{\mathrm{dir}}, R_{\mathrm{dir}})$, however, parameters for indirect dimensions are consistently determined and with enhanced precision. Since apparent noise level varies across direct dimension, the objective function is equipped with noise estimates:

$$Q(A,\omega,\mathbf{R}) = \sum_r \sigma_r^{-1} \sum_{\omega_m\in\Omega_r} (S_m^{\mathrm{residual}} - S_m^{\mathrm{model}})^2 \tag{3.7}$$

where the index r encompasses the direct dimension.

The initial parameters are $R = 0$, ω corresponding to the peak maximum, and the amplitude scaled appropriately to match the intensity at the maximum. These parameters are varied simultaneously to minimize target function (eq. 3.7) using Polak-Ribiere conjugate-gradient method (Press et al 2007, §10.8). The necessary analytical derivatives of the objective function can be evaluated with a limited overhead[1]:

$$\frac{\partial Q}{\partial A} = -2A^{-1}\sum_{\omega_m\in\Omega}(S_m^{\mathrm{res}} - S_m^{\mathrm{model}})S_m^{\mathrm{model}} \tag{3.8}$$

$$\frac{\partial Q}{\partial R} = -2\sum_{\omega_m\in\Omega}(S_m^{\mathrm{res}} - S_m^{\mathrm{model}})\frac{\partial S_m^{\mathrm{model}}}{\partial R} \tag{3.9}$$

$$\frac{\partial Q}{\partial \omega} = -2\sum_{\omega_m\in\Omega}(S_m^{\mathrm{res}} - S_m^{\mathrm{model}})\frac{\partial S_m^{\mathrm{model}}}{\partial \omega} \tag{3.10}$$

One can notice that the latter two derivatives can be evaluated using

$$\frac{\partial S^{\mathrm{model}}}{\partial R} = \mathrm{Re}\left\{\sum_{k=1}^{K}(-t_k)s(t_k;\ A,\omega,R)\exp(-i\omega_m t_k)\right\} \tag{3.11}$$

$$\frac{\partial S^{\mathrm{model}}}{\partial \omega} = \mathrm{Re}\left\{\sum_{k=1}^{K} i\,t_k s(t_k;\ A,\omega,R)\exp(-i\omega_m t_k)\right\} \tag{3.12}$$

[1]for brevity, the sum over direct dimension was omitted here

i.e. simply employing real-valued DFT of the synthetic signal (eq. 3.5), multiplied by $(-t_k)$ and $i\,t_k$, respectively. For performance reasons, analytical gradients are not evaluated using line-search subroutines, which rely entirely on function values.

In order to restrict variations of ω to the peak area, the following transformation of variables ($\omega = \omega(\theta)$) is used:

$$\begin{array}{rcl} \omega & = & \omega_{\min} + \pi^{-1}(\omega_{\max} - \omega_{\min})\,\mathrm{cot}^{-1}(-\theta) \\ \frac{\partial\omega}{\partial\theta} & = & \pi^{-1}(\omega_{\max} - \omega_{\min})(1+\theta^2)^{-1} \end{array} \tag{3.13}$$

which ensures that $\omega \in]\omega_{\min}, \omega_{\min}[$ for unconstrained parameter $\theta \in \mathcal{R}$. Decay rates are treated less rigorously: there is an additional penalty term $P(R) = \frac{2}{100}Q(R=0)\cdot R^2$ for $R < 0$, and a similar one for $R > R_{\max}$, which prefers decay rates to remain in the allowed range. The upper bound for R is inferred from the width of peak area ($\Delta\omega$), assuming that a maximally broadened Lorentzian line shape should decrease to a (somewhat arbitrary) relative intensity $t = 0.2$. This assumption, leading to $R_{\max} = \pi\Delta\omega\sqrt{t/(1-t)}$, is fairly non-restrictive (optimistic) since line shapes are usually broadened by additional factors, e.g. a decaying sampling density.

The model could be extended to flexibly describe more complex signals, e.g. modulated by unresolved scalar homonuclear couplings (^{13}C-^{13}C, ^{1}H-^{1}H). One possibility is to use a mixture of exponentially and Gaussian decay of the signal:

$$S(t;\ A,\omega,R,p,\sigma) = A\exp(i\omega t)(p\exp(-Rt) + (1-p)\exp(-\frac{1}{2}\sigma^2t^2)) \tag{3.14}$$

To avoid instability of parameters, we additionally assume equal line widths at half-height for Gaussian and Lorentzian line shapes, see Fig. 3.3. Effectively, $\sigma = R\,(2\,\ln 2)^{-1/2}$, and the only extra parameter is the relative weight p (assumed $p = 0.5$ as default).

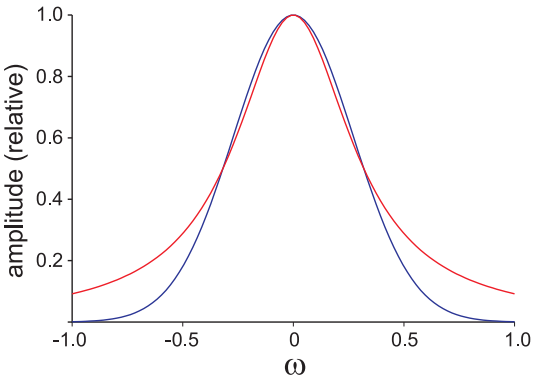


Fig. 3.3 Difference in line shapes between Gaussian (*blue*) and Lorentzian (*red curve*) functions. Equal maxima and line widths at half-height were assumed. Lorentzian decay rate $R = 2$ was used.
In general, Gaussian line shapes decay more rapidly than Lorentzian ones, and are thus more suitable for description of peaks containing unresolved splittings.

No effort is made to deconvolve overlapped peaks since it is prone to numerical unstability. Also, by no means would it represent a general solution for crowded spectra of biomolecules. Clearly, for some peaks the assumed simple model may be inappropriate. A possible criterion for acceptance of obtained fit is goodness-of-fit (reduced χ^2) but it suffers from incertainty of independent variables that should determine χ^2 statistic. Instead, a more practical solution is to compare denoising performance of analytic and raster model. Let us assume that after subtraction of FID simulated from analytical and raster model the apparent mean noise level in the intersecting indirect (N-1)-dimensional spectra are σ_a and σ_d, respectively. Then the probability of acceptance of analytical model is

$$P = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\sigma_r - \sigma_a}{\sqrt{2}\cdot 0.01\bar{\sigma}}\right) \tag{3.15}$$

where $\bar{\sigma} = \sqrt{\sigma_a^2 + \sigma_r^2}$, and the denominator determines the magnitude of meaningful difference between σ_a and σ_r. Effectively, $P = \frac{1}{2}$ if both methods provide equal noise levels in the residual spectrum, $P \to 1$ if $\sigma_r \gg \sigma_a$, and $P \to 0$ in the opposite case.

3.1.5. Iterative model refinement

As mentioned in the sec. 3.1.1, all peak models should be optimized simultaneously due to mutual interactions via NUS artefacts. However, peaks are modelled sequentially for performance reasons. To ensure that all models are ultimately optimized in the same conditions (i.e. an equal noise level), the following *refinement* procedure is proposed:

1. Generate and sparsify the time-domain signal corresponding to the peak model
2. Reintroduce it to the residual FID
3. Recompute noise level in the relevant spectral region
4. Verify whether the putative peak still fulfills the intensity criterium
5. If so: determine again the peak area, i.e. expand it starting from the peak maximum
6. Update peak model (either raster or analytical)
7. Subtract the corresponding time-domain signal from residual FID

Note that only peaks intersecting (N-1)-subspectra where new peaks were found need to be refined. Otherwise, the artefact level remains similar and peak models cannot be improved. Previously determined peak models are used as starting point for refinement if they fit better than default initial ones.

As a result, the particular peaks are modelled *as if other ones were already removed* from residual spectrum (or FID). It has been shown to improve linearity of peak amplitudes, and also, to a certain extent, prevent from introduction of false positives. Such features should roughly cancel with artefact interferences of opposite sign after subtraction of peaks they originate from.

3.1.6. Spectral reconstruction

Reconstruction of spectra *as if they were* uniformly sampled is the ultimate goal of the algorithm. The meta-data generated by SSA contain information about the systematic portion of input signal: fitted peak parameters (amplitude, decay rates, frequencies) and N-dimensional spectral regions (peak models g) for overlapped or non-Lorentzian peaks. Reconstruction should also incorporate the „tiny", but sometimes crucial information left in residual sparse FID, the second output of SSA (see Fig. 3.4). The thermal noise remaining in the residual spectrum is also necessary to assess relevance (e.g. in terms of S/N) of spectral features in the reconstruction.

The residual *sparse* signal is processed with z.-a. FT, whereas peak models require first the generation of *full* time-domain signal. The latter is then processed with conventional DFT to yield "clean" part of the spectrum. Let us comment that it also contains a certain amount of localized noise, which is transferred as a result of raster model adjustment or slightly imprecise fitting of analytic model.

The direct addition of both parts would imply artificial peak amplification, or nonlinearity, since peak models are simulated on a *full* but residual FID on *sparse* time-domain grid. To cope with this imbalance, every peak model is simulated on

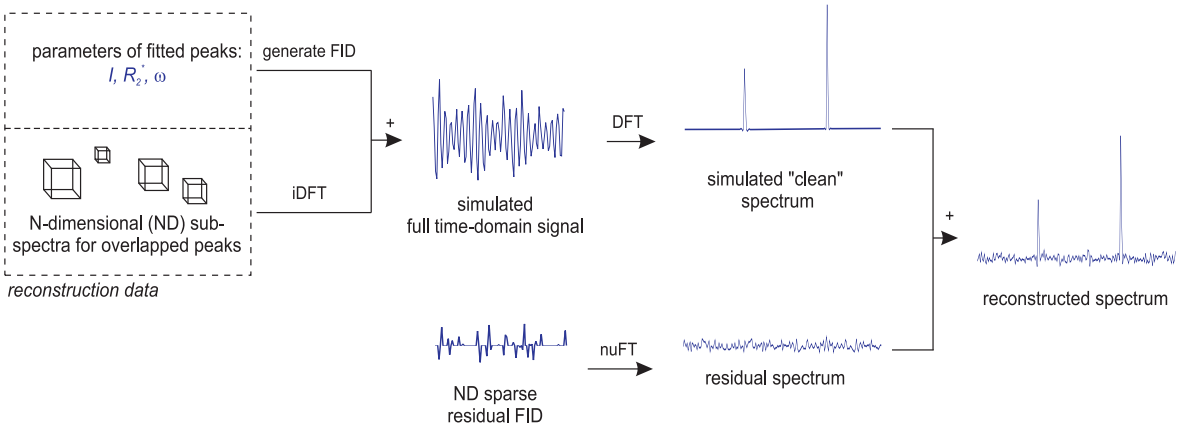


Fig. 3.4 The conceptual scheme of spectral reconstruction from meta-data and residual FID generated by SSA.

1. a sparse experimental grid, and Fourier transformed (z.-a. FT),
2. and on a full grid, and Fourier transformed (DFT).

The ratio of peak maxima in both cases is then used to down-scale the full time-domain synthetic signal. The described procedure need not be used in reconstruction algorithms which model all data on a full time or spectral grid, and discard any residuals, e.g. MaxEnt, CS, MDD (see sec. 2.4). However, non-linear modelling of noise and weak peaks is challenging, and may lead to troublesome distortions.

Note that it is possible to reconstruct spectral data with increased resolution and additional apodization of time-domain data, similarly to how it can be manipulated prior to conventional DFT.

3.2. Implementation and performance issues

3.2.1. Multidimensional extensions

Extension of all above described procedures to a multidimensional case poses no conceptual difficulty. In particular, a general real-valued z.-a. FT of *on-grid* data is employed (Kazimierczuk et al 2012), using sequential approach as justified in sec. 3.2.3. Up to now the algorithm has been implemented for 2D-5D data. To emphasise the uniformity of the approach since 2011 the program was reimplemented to maximally unify the source code. This was accomplished with the substantial contribution of **Krzysztof Kosiński**, a former M. Sc. student co-supervised by the author. The generalized data structures were designed with the extensive use of generic programming features of C++ language, namely recursive *dimensionality-dependent* C++ templates. At present the only dimensionality-specific code includes low-level FT transformation procedures.

However, there is a significant difference in peak searching routine for 5D case. In general, exploring the entire and vastly empty 5D spectrum is usually an achievable, but unnecessarily slow process. Instead, we resorted to the idea of sparse MFT (Kazimierczuk et al 2009, 2010). Accordingly, some of peak coordinates can be inferred from a lower-dimensional spectrum, such as 2D HSQC or 3D HNCO. As shown in Fig. 3.6, the peak searching in 5D spectrum starts from the spectral blocks intersecting the pre-defined 2D cross-sections. Their positions are determined by entries in a peak list from 3D *base spectrum*. Note that peak areas can

be extended outside these blocks. However, if the 3D peak lists are incomplete or highly inaccurate, some sufficiently intense 5D peaks may be not found in the restricted search area. Effectively, denoising performance of 5D SSA can be affected. Other aspects such as peak modelling using N-dimensional raster or analytical models remain unchanged. It should be emphasised that to ensure reasonable performance and memory consumption up to 5D case, specific optimizations, described in the next sections, were necessary.

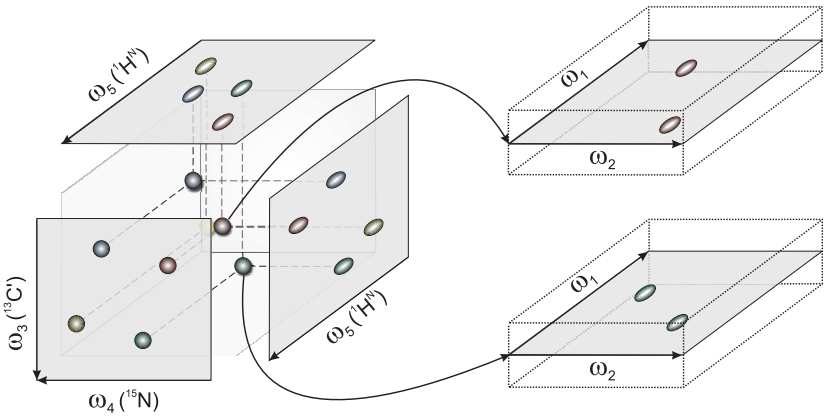


Fig. 3.5 Peak search area in 5D SSA is determined by predefined two-dimensional cross-sections, computed on a basis of peak positions in 3D *base* spectrum, e.g. HNCO. The vicinity of cross-sections is also explored. Effectively, the approach is tolerant to small errors in peak list entries.

The implementation of 3D and 4D SSA is available at no charge for academic purposes on the web site of our group: nmr.cent3.uw.edu.pl. A complete package consists of binaries `cleanerNd` and `reconstructorNd` for processing and reconstruction, respectively, and makes use of external program(s) `toastd`, and optionally `nmrPipe` (Delaglio et al 1995), for spectrometer data conversion. Currently, the 5D version is only available upon request.

3.2.2. Memory management

SSA has been tailored for challenging spectra of very high resolution and dimensionality. Full spectral matrices of 3D spectra range to a few GB, and cannot be entirely stored in machine's memory (RAM). In the 4D case, even a 3D subspectrum may excess the capacity of RAM, and a full spectrum can occupy vast amounts of disk space (up to hundreds GB). Finally, 5D spectra are unlikely to be stored in neither RAM or permanent disk memory.

The only general solution of this problem is to divide spectral domain into blocks, preferably of fixed size, cf. Fig. 3.6. The usual choice of block size (`BS`) is 32, 16 and 8 for 3D, 4D and 5D case, respectively. Effectively, the memory consumption of a single block in the single precision is $m = (BS)^{D-1} \times 4\text{B}$, corresponding to 4 kB, 16 kB, and 16 kB for 3D, 4D and

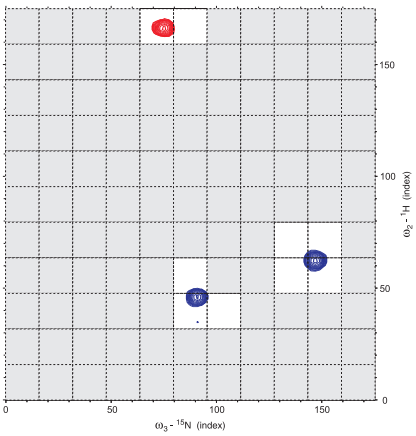


Fig. 3.6 Division into blocks of size 16 × 16 shown on a 2D cross-section of 4D HN(CA)NH spectrum. High-dimensional spectra of biomolecules are usually quite sparse, therefore storage and processing of selected spectral blocks, instead of full matrices, yields significant speed-up and memory savings.

5D blocks, respectively. The optimal block size, being the trade-off between performance of DFT processing routines and overall memory consumption, was determined by **K. Kosiński** in his M. Sc. thesis (2013). Additionally, to account for the incomplete occupation of multidimensional blocks, compression using deflate-inflate algorithms available in `zlib` library (zlib.net) were employed for peak models not processed at the given time.

Note that the direct dimension is not divided into blocks (or equivalently `BS`=1) since this would only increase memory consumption with little effect on processing performance. In addition, the author designed continuous memory data structures for storage and manipulation of sparse and mixed sparse-regular data arrays which occur as intermediates during sequential z.-a. FT. Dedicated sparse data structures for storing information on peak areas, which utilize *hash tables* and *lists*, are described in M. Sc. thesis of Kosiński (2013).

3.2.3. Optimization of DFT core

Processing performance was of the particular concern in implementation of SSA. Below, we explain that in multiple dimensions the restriction to on-grid sampling enables numerical optimization of two kinds.

Let us consider the two-dimensional case, with typical sparse distribution of N samples shown in Fig. 3.7. If time-domain data are distributed *off-grid*, the total computational cost of real-valued FT amounts to $M_1 M_2 N$, where M_i are digital resolutions. At each transformation stage (dimension), the cost increases by a factor of M_i since time coordinates in remaining dimensions are generally unique (see bottom panels in Fig. 3.7). In contrast, if number of samples N distributed *on a* $N_1 \times N_2$ *grid* is larger than number of increments in any of the indirectly sampled dimensions, $N > N_1$, $N > N_2$, ..., there is large probability of clustering at the subsequent processing stages. E.g., there are only 9 instead of 25 unique t_2 indices for the second transformation considered in Fig. 3.7. The transformations require $M_1 M_2 N_2 +$

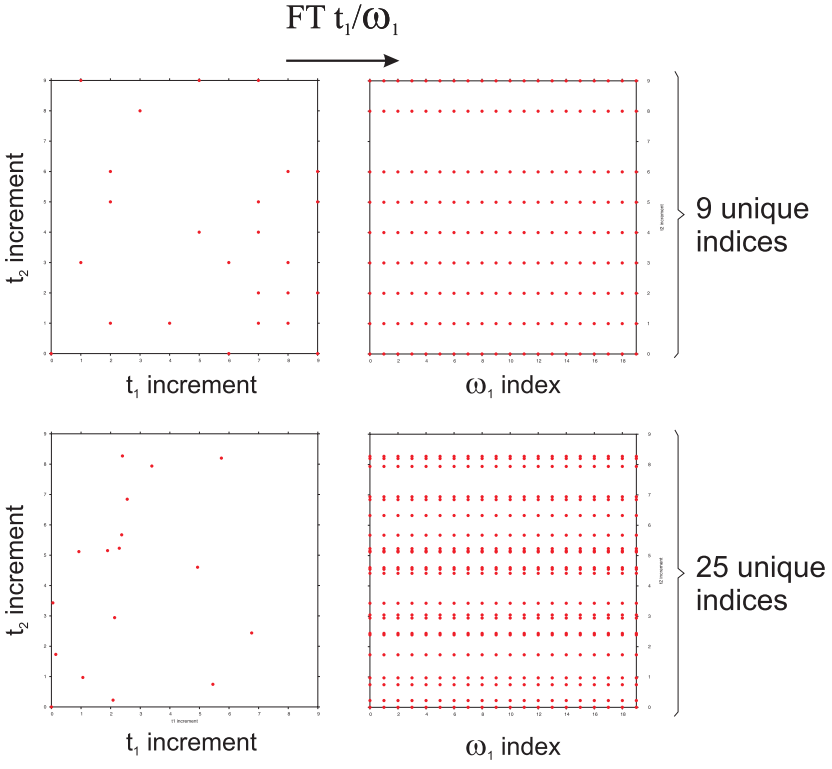


Fig. 3.7 Performance advantage of on-grid point distribution shown in the two-dimensional case. Due to coincidence of t_2 coordinates (*top left*) FT $t_1 \to \omega_1$ generates reduced number of rows, which are subject to FT in t_2/ω_2 dimension. In contrast, uniqueness of t_2 coordinates in the arbitrary *off-grid* distribution (*bottom left*) results in an unchanged number of rows ($\approx N$) for the subsequent transformation stages.

$M_1 N$ operations in total, which is approximately a factor of N/N_2 less than for arbitrary off-grid distribution. It should be emphasised that these speed-up factors increase with dimensionality since (usually) more points N are used, whereas $\min(N_i)$ remains unchanged. The analysis also suggests that it is beneficial to reorder dimensions so to process the one with the least number of increments at the last stage.

Noteworthy, at every stage of transformation of *on-grid* data, the data density significantly increases as shown in Fig. 3.8. As already explained in sec. 2.2.2, under such circumstances FFT may perform competitively to "slow" DFT. As the actual performance depends highly on the schedule, hardware and libraries used, the SSA implementation adopts the mixed strategy: the empirically faster of FFT and DFT routines are used.

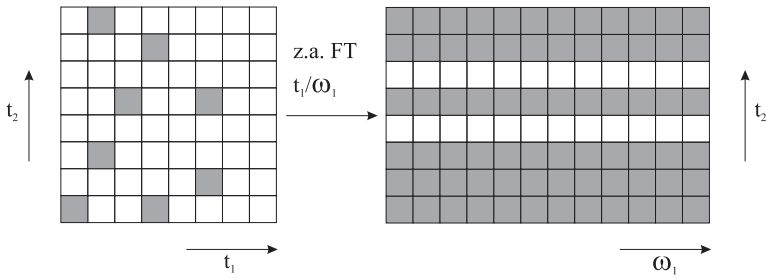


Fig. 3.8 Difference in data sparseness at different stages of Fourier processing of multidimensional NUS data. While $t_1 \to \omega_1$ 1D transformations are generally quite sparse, the $t_2 \to \omega_2$ processing is fairly dense.

The DFT routines are implemented simply as linear algebra operations, mostly including matrix-vector (MV) and matrix-matrix (MM) operations. Fourier matrices are precomputed and kept permanently in machine's memory. For MV and MM multiplications SSA implementation utilizes basic linear algebra subroutines (BLAS) with FORTRAN API (Blackford et al 2002). The performance is further increased by use of vendor-specific implementation of BLAS, such as Intel Math Kernel Library or AMD Core Math Library. FFTs are performed with the cutting-edge algorithms implemented in the "Fastest Fourier Transform in the West" library (Frigo and Johnson 2005).

SSA program also benefits from the common accessibility of multi-core and multi-processor systems. The evident parallelism of either linear algebra operations, or possibility to distribute computation of indirect blocks (or division along the direct dimension) among independent threads, is utilized. A well-standardized OpenMP library is employed for parallel computing at shared memory machines.

3.2.4. Efficient spectral reconstruction

The typical size of the *reconstructed* time-domain can be roughly estimated to $(10^2)^D$ points. This clearly shows that its storage and/or computation may be quite impractical or even impossible yet at dimensionality D=4. To cope with this difficulty, reconstruction data are always stored in their compressed form (see Fig. 3.4). Due to limited size it can be transferred from a high-performance server to another machine, and reconstructed there. Also, the output of SSA can be used multiple times for reconstruction of different spectral regions, or with modified resolution. For these reasons, the reconstruction procedures were isolated from the main SSA iterations, and implemented in a separate program `reconstructorNd`. In conclusion, the workflow is as follows:

$$\text{NUS FID} \xrightarrow{\text{cleanerNd}} \text{reconstruction data} \xrightarrow{\text{reconstructorNd}} \begin{cases} \text{ND spectrum (or its subregion)} \\ \text{2D cross-sections} \\ \text{2D projections} \end{cases}$$

Quite importantly, due to prohibitive size a full time-domain signal should **not** appear as an intermediate in the reconstruction of spectral ND-blocks or 2D cross-sections:

$$\text{raster ND-model} \xrightarrow{\text{iDFT}} \textbf{full-grid FID} \xrightarrow{\text{DFT}} \text{ND-block or 2D cross-section}$$

The remedy is to interleave sequential DFT/iDFT operations, which can be performed in arbitrary order without affecting the result. E.g., instead of the usual sequence in a 4D case:

$$\begin{array}{rll} \text{raster ND-model} & \xrightarrow{\text{iDFT}(t_1)}\ \xrightarrow{\text{iDFT}(t_2)}\ \xrightarrow{\text{iDFT}(t_3)} & \textbf{full-grid FID} \\ & \xrightarrow{\text{DFT}(t_3)}\ \xrightarrow{\text{DFT}(t_2)}\ \xrightarrow{\text{DFT}(t_1)} & \text{ND-block} \end{array}$$

the same result can be obtained in the sequence

$$\text{raster ND-model} \xrightarrow{\text{iDFT}(t_1)} \xrightarrow{\text{DFT}(t_1)} \xrightarrow{\text{iDFT}(t_2)} \xrightarrow{\text{DFT}(t_2)} \xrightarrow{\text{iDFT}(t_3)} \xrightarrow{\text{DFT}(t_3)} \text{ND-block}$$

where required intermediate storage is only slightly larger than the size of input model. The contribution from analytic models are simply computed as the direct products of 1D-shapes, and requires little computational time. Noteworthy, fully-dimensional spectra can be reconstructed in the used-defined restricted area. This task can also be distributed to multiple parallel threads. Computation is performed in a block-by-block manner, while the disk write overhead is reduced by use asynchronous I/O operations. In addition, both ND spectra and projections are updated on-fly and can be inspected during (sometimes) time-consuming reconstruction. As a curiosity, the interleaved scheme of transformations and block-approach allow to compute projections from fully dimensional 5D spectrum, even if the latter cannot be stored using conventional disks!

Finally, let us comment on the computation of 2D cross-sections, which might be useful for certain 4D spectra, and represents the only reasonable approach to inspection of 5D spectra (Kazimierczuk et al 2009). Let us consider the $\omega_1 \times \omega_2$ cross-sections which are computed employing $\omega_3, \ldots, \omega_N$ coordinates (assumed to be known in advance). The optimal order of transformation is

$$\begin{array}{rll} \text{raster ND-model} & \xrightarrow{\text{iDFT}(t_N)}\ \xrightarrow{\textbf{sDFT}(t_N)} & \\ & \cdots & \\ & \xrightarrow{\text{iDFT}(t_3)}\ \xrightarrow{\textbf{sDFT}(t_3)} & \\ & \xrightarrow{\text{iDFT}(t_2)}\ \xrightarrow{\text{DFT}(t_2)} & \\ & \xrightarrow{\text{iDFT}(t_1)}\ \xrightarrow{\text{DFT}(t_1)} & \text{ND-block} \end{array}$$

in which $D-2$ sparse FTs (**sDFT**), i.e. to a single point in the frequency domain, precede two final full-dimensional $t_2 \to \omega_2$, and $t_1 \to \omega_1$ transformations. Clearly, the similar conclusion can be drawn for transformation of sparse residual FID (Stanek et al 2012). Note that the optimal computation of different cross-sections, e.g. $\omega_1 \times \omega_2$ or $\omega_2 \times \omega_3$, requires the flexibility of internal dimension reordering in `reconstructorNd`. To simplify analysis of cross-sections, they are merged into one pseudo-3D spectrum, which can be conveniently inspected in `Sparky` program[2] (Goddart and Kneller).

[2]The original solution of Kazimierczuk et al (2009) was to store every cross-section in a separate file, which proved inconvenient in navigation and analysis

3.3. Discussion and comparisons

The performance of the algorithm was first tested on synthetic signals, and compared with CLEAN in the version implemented by Coggins and Zhou (2008). In the first case, the denoising efficiency was investigated on a signal containing a single dominant component. A second small peak of relative intensity of 3% was added as a probe for spectral quality. Gaussian noise of σ=0.01 (relative to the amplitude of the main signal) was also added. The signal was sampled randomly with 75 points, corresponding to 15 % of Nyquist grid. The results of z.-a. FT, CLEAN and both *raster* and *analytic* model fitting are shown in Fig. 3.9.

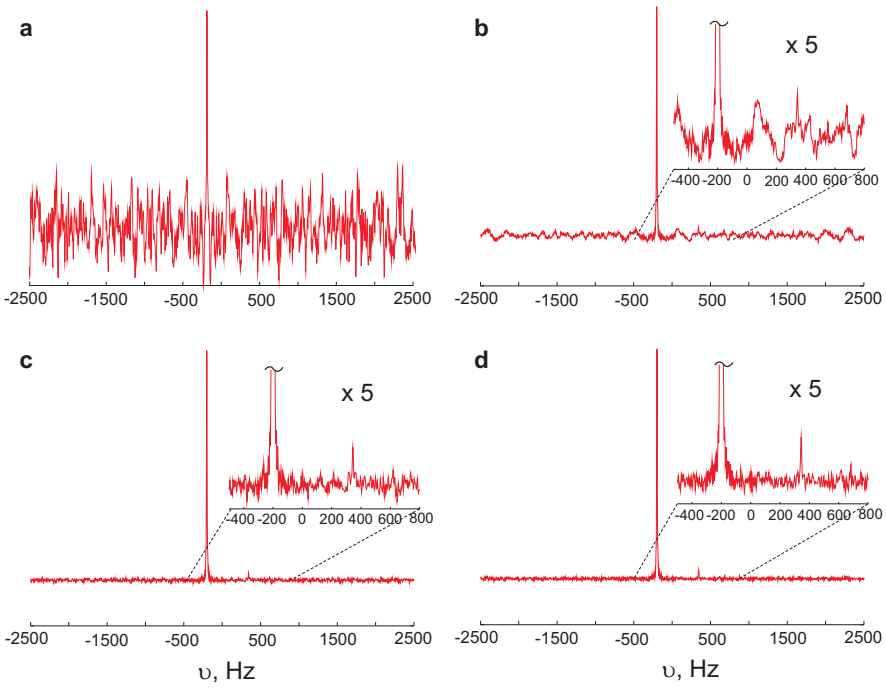


Fig. 3.9 The comparison of results of z.-a. FT (*a*), CLEAN (*b*) and *raster* (*c*) and *analytic* model fitting (*d*) in SSAlgorithm. Two components of equal relaxation rates of R=25 Hz, and relative intensities of 1 and 0.03 were generated. Spectral widths of 5 kHz and max. evolution time of 10 ms were assumed. 75 sampling points were used. The insets show the magnified spectral region containing the smaller spectral component.

The zero-augmented FT performs inferior to other methods yielding the spectrum of the poor S/N ratio. CLEAN was capable of removing majority of NUS pseudo-noise, however, the tiny peak is still obscured by residual artefacts. In contrast, both approaches to signal modelling incorporated in SSAlgorithm suppressed artefacts sufficiently well to enable detection of the small resonance. This very simple example shows that SSA ensures a more efficient restoration of missing samples. This can be attributed to the fact that SSA models a complete peak, while CLEAN builds it gradually (depending on the loop gain parameter) and leaves some signal in the residual spectrum as soon as its intensity reaches the peak detection threshold.

The second synthetic signal contained multiple components of significantly different intensity of the ratio 32:16:8:4:2:1 (Fig. 3.10). Different amount of Gaussian noise was added. Again, z.-a. FT resulted in significant loss of information since only two strongest peaks were above NUS pseudo-noise. CLEAN performed slightly better, enabling detection of the third

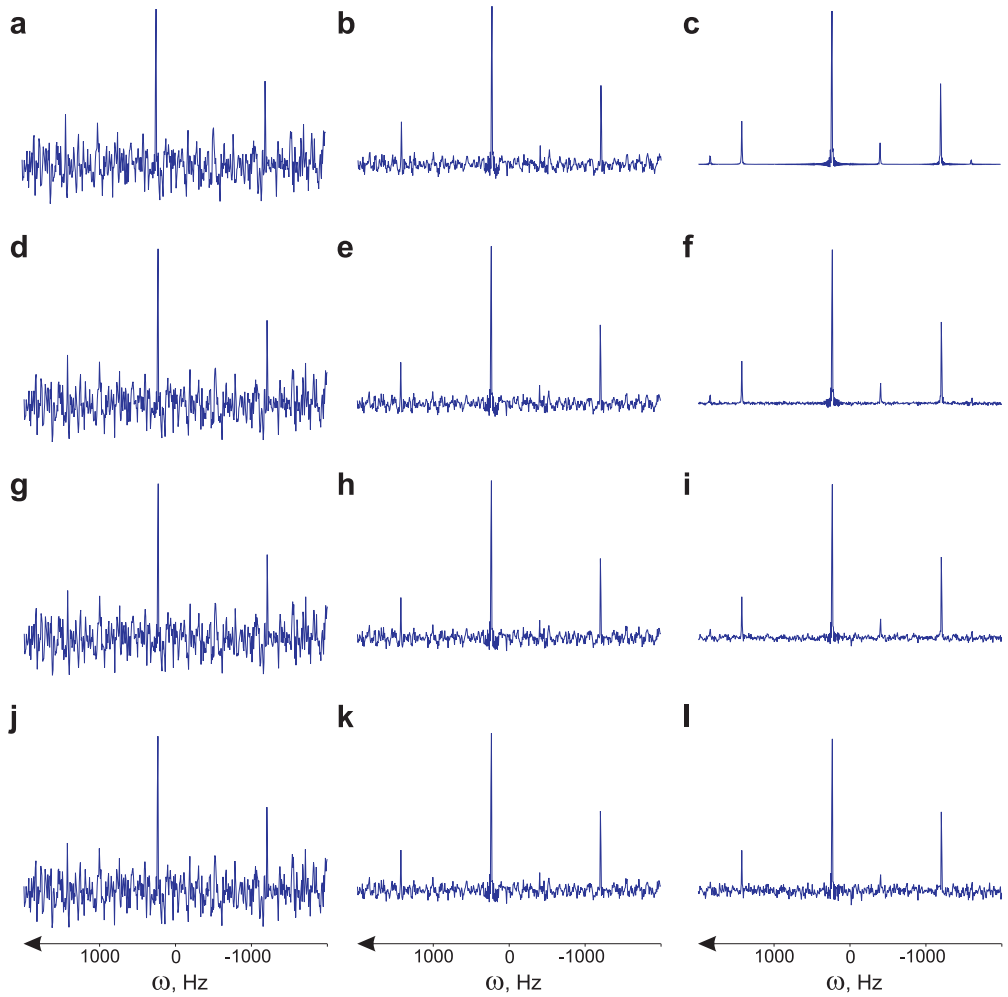


Fig. 3.10 l_2-norm (*a, d, g, j*), CLEAN (*b, e, h, k*) and SSA (*c, f, i, j*) reconstructions of a synthetic signal sampled at 70 out of 280 equidistant points. Six components of equal relaxation rate and relative amplitudes of 1:2:4:8:16:32 were simulated. The increasing amount of Gaussian noise was added (from *top* to *bottom*): $\sigma = 0$ (*a-c*), 0.02 (*d-f*), 0.04 (*g-i*) and 0.08 (*j-l*), where 1 refers to the amplitude of the strongest component.

component. However, even for the noiseless signal (Fig. 3.10*b*) the reconstruction was quite imperfect, showing a significant contribution of residual artefacts. Analytic model fitting restored all available information in this case, revealing even the smallest component. Despite noiseless signals do not occur in practice, this test shows the limiting capabilities of CLEAN and SSA. In this case, the crucial advantage of the latter is the use of the proper signal model. Also, SSA avoids error accumulation present in CLEAN, and thanks to the refinement of peak models the correct intensities of all resonances are reproduced.

In the following, the SSAlgorithm was compared to two competitive approaches: MDD and MaxEnt, using experimental data. Automatic parameter estimation for MaxEnt was used to avoid any bias (Mobli et al 2007). For MDD, the number of expected components was provided as estimated from signal count in the respective region on ^{15}N-HSQC of human ubiquitin. In the first test, a severly undersampled 3D HNCA spectrum was investigated. As can be inferred from Fig. 3.11, the spectrum is affected by NUS artefacts and some weaker peaks are at the edge of detection after z.-a. FT processing. Other employed methods yield comparable results, restoring all relevant spectral information. Non-linear effects are pro-

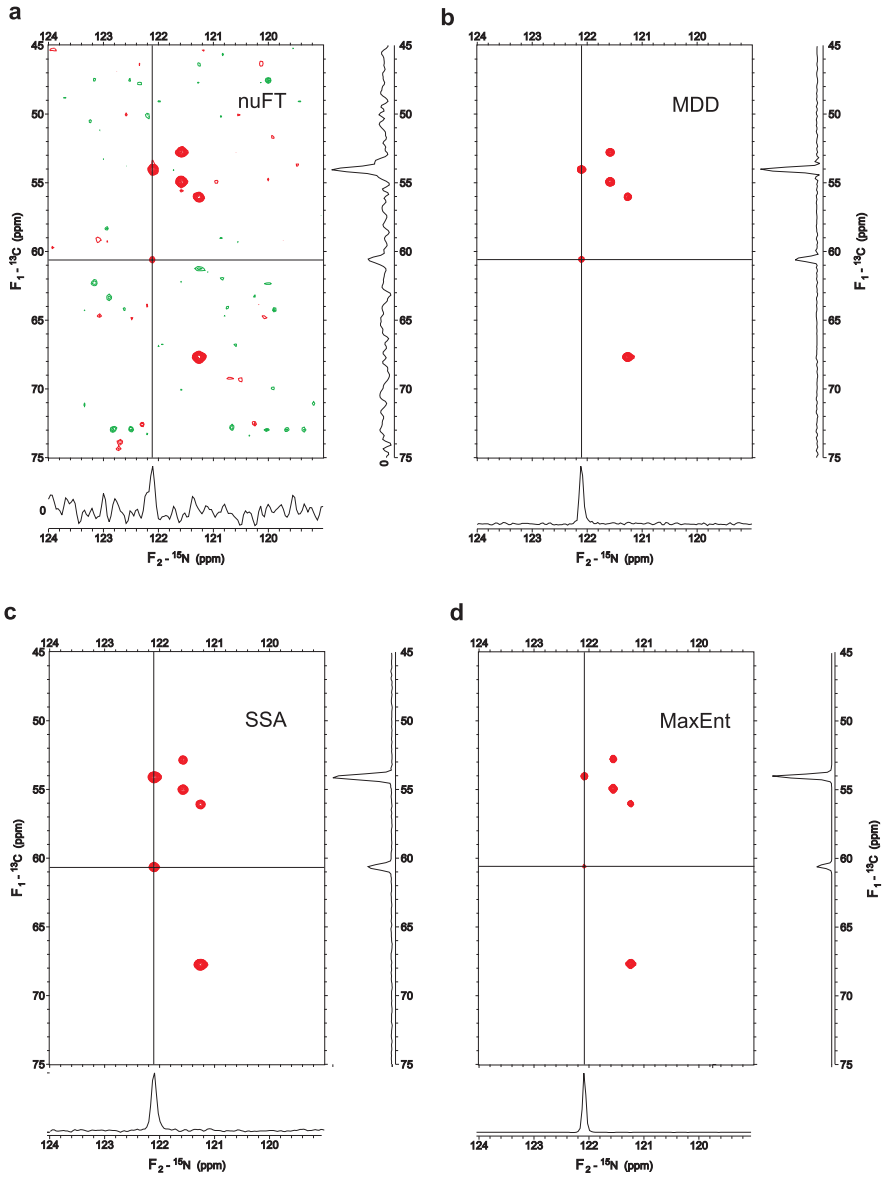


Fig. 3.11 ω_1 (C^α) – ω_2 (^{15}N) cross-sections of 3D HNCA spectra of human ubiquitin (at ω_3=8.112 ppm), obtained from the same NUS data with different processing methods: z.-a. FT (*a*), MDD (*b*), SSA (*c*), and automatic MaxEnt (*d*). The cross-sections are plotted with the same intensity scale. 1024 signal samples were used. Max. evolution times of 12.9 and 160 ms were achieved in C^α and ^{15}N dimensions, respectively. Relative sampling density $\theta = 0.032$ was obtained.

nounced for the MaxEnt case, where noise intensity is underestimated. Noteworthy, despite quite low *relative* sampling density, MDD performs well. The number of samples seems sufficient to converge MDD model with relatively sparse signals (compare eq. 2.75).

The next test was performed using high-resolution 3D NOESY-^{15}N-HSQC, randomly sampled at very low level (θ=0.027). The results of MDD, MaxEnt and SSA are compared with the low-resolution time-equivalent conventional spectrum in Fig. 3.12. Due to non-linear transformation of peak amplitudes, using the same lowest contour MaxEnt spectrum displays less peaks than other methods. Not necessarily the weakest peaks are missing, however, highly incorrect peak volumes hinder distance calibration. Alternatively, peak intensities can be corrected using *in silico* calibration (Schmieder et al 1997). MDD reconstruction shows

excessive number of peaks, particularly at the crossings of shapes of the strongest components. In fact, the noise level in the *shapes* is somewhat elevated w. r. t. remaining regions of the spectrum. The difficulty in convergence of MDD model for highly undersampled data is in agreement with Orekhov's simulations and can be justified with insufficient data compared to the number of variables in MDD model (Orekhov et al 2003). At these conditions, SSA spectrum shows the best agreement with the conventional Fourier spectrum.

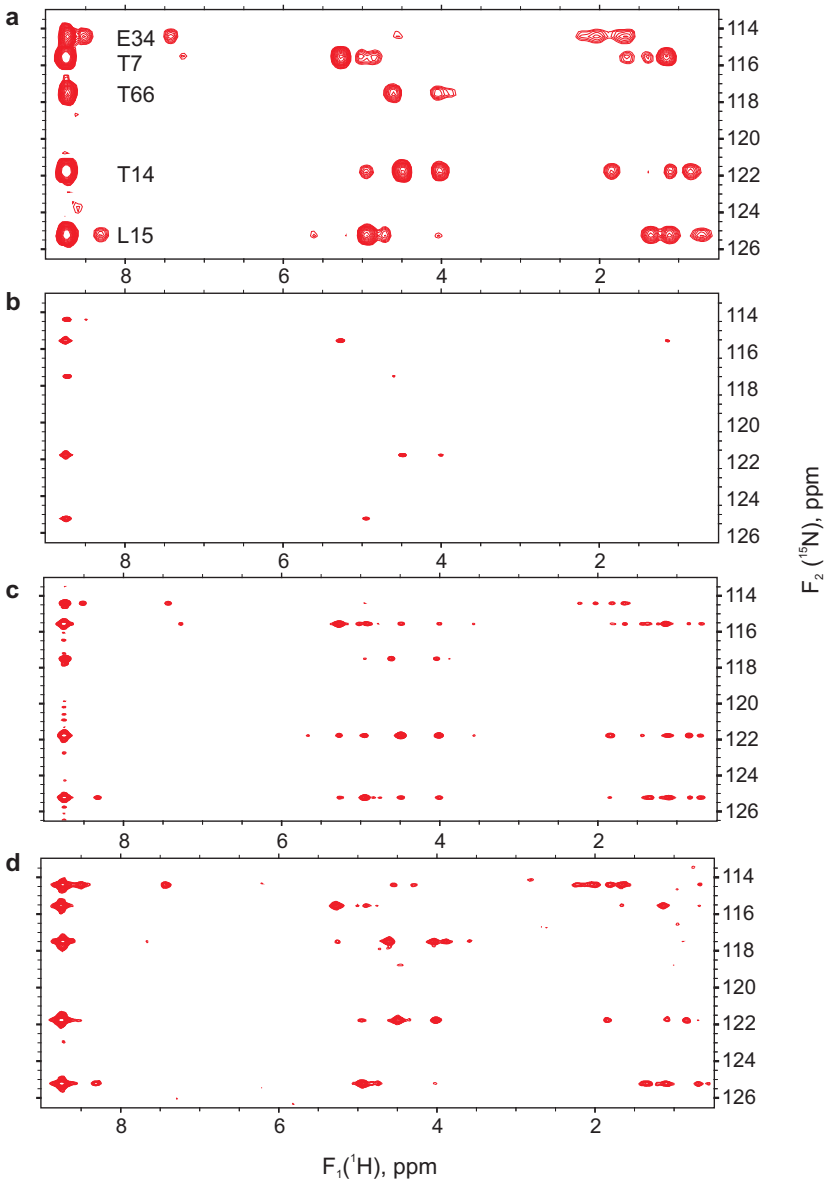


Fig. 3.12 ω_1 (^{1}H) – ω_2 (^{15}N) cross-sections of 3D NOESY-^{15}N-HSQC spectra of human ubiquitin at ω_3=8.743 ppm: (*a*) the conventional spectrum, (*b-d*) spectra from the same NUS data obtained with different methods: MaxEnt (*b*), MDD (*c*) and SSA (*d*). The cross-sections are plotted using the same intensity scale (cutoff level at 1.5 % of T14 diagonal peak). Maximum evolution times of 30 and 100 ms and spectral widths of 12 and 2.4 kHz were set in ^{1}H and ^{15}N dimensions, respectively. 2300 sampling points were used, yielding relative density of $\theta = 0.027$.

3.4. Evaluation and benchmarks

Linearity Correct reproduction of relative peak amplitudes is an important and desired feature. It is of particular relevance in NOESY or relaxation applications, where peak heights (or volumes) carry quantitative information. SSA was tested in this respect on NUS 3D NOESY-^{15}N-HSQC spectrum. The results were compared to a fully sampled reference spectrum as described in detail in sec. 4.2.2 on page 98. Peak peaking and manual integration were performed as in the usual protocol of analysis. The overlapped peaks with diverging Gaussian fits were discarded to avoid systematic errors. The comparison of peak volumes showed an excellent linear correlation over two orders of magnitude. The correlation coefficient R^2=0.998 was obtained.

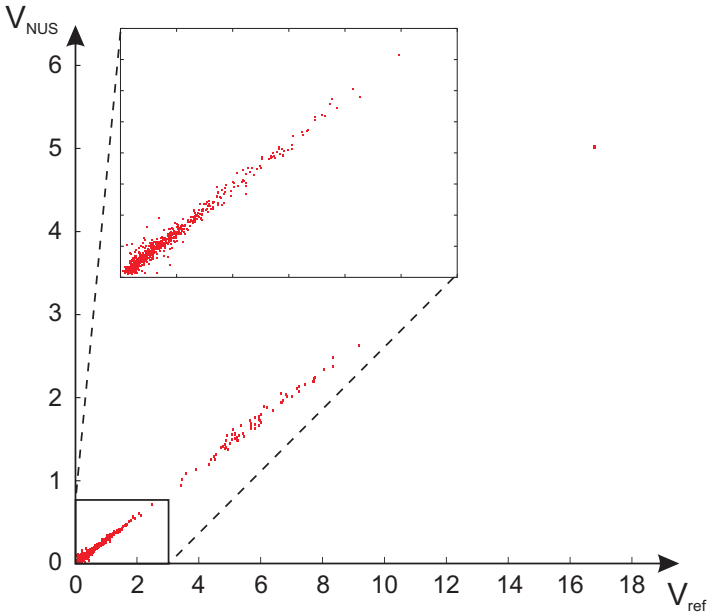


Fig. 3.13 Correlation plot of peak volumes (in arbitrary units) of diagonal and crosspeaks for 70 out of 76 backbone residues of human ubiquitin (except M1, P19, E24, P37, P38 and G53), measured in randomly sampled (V_{NUS}) and reference ^{15}N-edited NOESY-HSQC spectra of the same resolution (V_{ref}). Peak volumes were calculated with SPARKY utility using Gaussian fit; procedure converged in both spectra for 732 peaks (752 and 769 in reference and randomly sampled ones, respectively).

Denoising performance Another crucial feature of the reconstruction algorithm is its capability of NUS artefact suppression in application to real NMR signals. In practice, phase distortions, spurious dispersive peaks and unevenly distributed noise may affect the denoising performance of the algorithm. Ideally, to focus on artefact suppression only, a spectrum with the lowest influence of thermal noise is preferred. Therefore, the most meaningful results are obtained when SSA is applied to very *sensitive* spectra such as 4D HCCH-TOCSY performed on a sample of human ubiquitin in decent concentration (1 mM). Other suitable examples are 3D ^{13}C-edited and 4D ^{13}C,^{13}C-edited NOESY spectra of ubiquitin, discussed in sections 4.12 (on page 99) and 4.17 (page 104).

To quantify the performance of SSAlgorithm, the new measure was introduced. The level of noise reduction, ρ_n, is defined as:

$$\rho_n = \frac{\sigma_0 - \sigma}{\sigma_0} \tag{3.16}$$

where σ and σ_0 are the noise levels in SSA and z.-a. FT spectra, respectively. Due to unavoidable presence of thermal noise, this parameter can never reach the unity, however, for the sensitive experiments its influence can be safely neglected. The results for 4D HCCH-TOCSY are shown in Fig. 3.14. The noise reduction varies across the directly detected

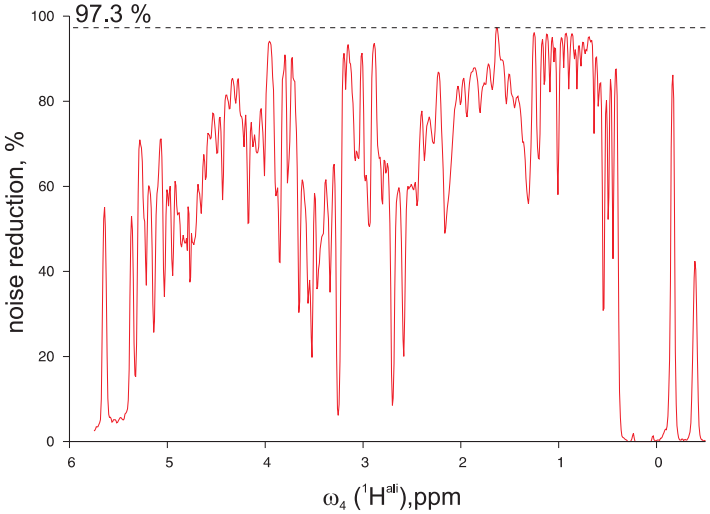


Fig. 3.14 Noise reduction (defined by eq. 3.16) in NUS 4D HCCH-TOCSY spectrum of human ubiuitin. The noise is measured as described in section 3.1.2.

dimension, depending on the signal strength. In the most intense methyl region (0.5-1.5 ppm) the apparent denoising limit is reached at $\rho \approx 97\%$. In the case of two aforementioned NOESY spectra, the respective values of $\rho \approx 98\%$ are obtained. In other words, in the case of purely *sampling limited* spectra, SSA yields approximately 50-fold artefact reduction or, equivalently, 50-fold sensitivity enhancement. When applied to typical NMR data with non-negligible contribution of thermal noise, smaller gains are expected. Nevertheless, SSA is beneficial also in these cases as demonstrated on numerous examples in the next chapter.

Chapter 4

Relevant experimental applications

4.1. Experiments for sidechain resonance assignment

In contrast to well-established backbone assignment protocols, the full assignment of sidechain resonances remains a substantial bottleneck in protein studies at atomic resolution using NMR spectroscopy (Hiller et al 2008b). The task is particularly difficult for „large" proteins, which backbone resonances can usually be assigned using ^{2}H-labelling and TROSY versions of triple-resonance techniques, but these sensitivity enhancements cannot be transferred to side chain assignment experiments. However, only rough global folds of large biomolecules can be established basing on H^N-H^N through-space contacts. Clearly, the complete list of side-chain ^{1}H resonances is a prerequisite for precise structure calculation based on NOE distance constraints. Nowadays, most commonly *Combined Automated NOE Assignment and Structure Determination* (CANDID) approach is employed for this purpose (Herrmann et al 2002). It has been demonstrated that these algorithms are quite prone to incomplete side chain assignment, and usual tolerance is fewer than 10% (Güntert 2004; Jee and Güntert 2003). In certain cases, such as essential methyl or aromatic resonances (buried in the hydrophobic cores of proteins and usually giving a number of NOE cross-peaks), a very few missing or incorrect assignments can lead to severe structure distortions. This is due to the fact that CANDID tends to find a wrong assignment of highly ambiguous NOE cross-peaks. Therefore, considerable efforts have been made to simplify and increase the robustness of sidechain assignment. The use of high-dimensional spectra with NUS for resolution enhancement are essential part of these advancements and will be commented in the following sections.

4.1.1. 4D HCCH-TOCSY (aliphatic spin systems)

HCCH-TOCSY experiment, proposed as 3D HC(C)H correlation by Bax et al (1990b), was later significantly improved by use of PFGs by Kay et al (1993). The 4D version which was soon published (Olejniczak et al 1992), fully utilized resolving power of all nuclei involved in the coherence transfer:

$$^{1}\mathrm{H}(t_1) \xrightarrow{^{1}J_{\mathrm{CH}}} {}^{13}\mathrm{C}(t_2) \xrightarrow{^{1}J_{\mathrm{CC}}} {}^{13}\mathrm{C}(t_3) \xrightarrow{^{1}J_{\mathrm{CH}}} {}^{1}\mathrm{H}(t_4)$$

The pulse sequence shown in Fig. 4.1 is fairly short, and utilizes most coherence transfer delays for shared-time evolution. In effect, this very sensitive experiment is routinely used for sidechain assignment of stably folded proteins. Nevertheless, *uniformly sampled* experiments are mostly performed without ^{1}H (t_1) evolution, causing common difficulties with correct matching of ^{1}H and ^{13}C chemical shifts. The potential utility of application of NUS was first

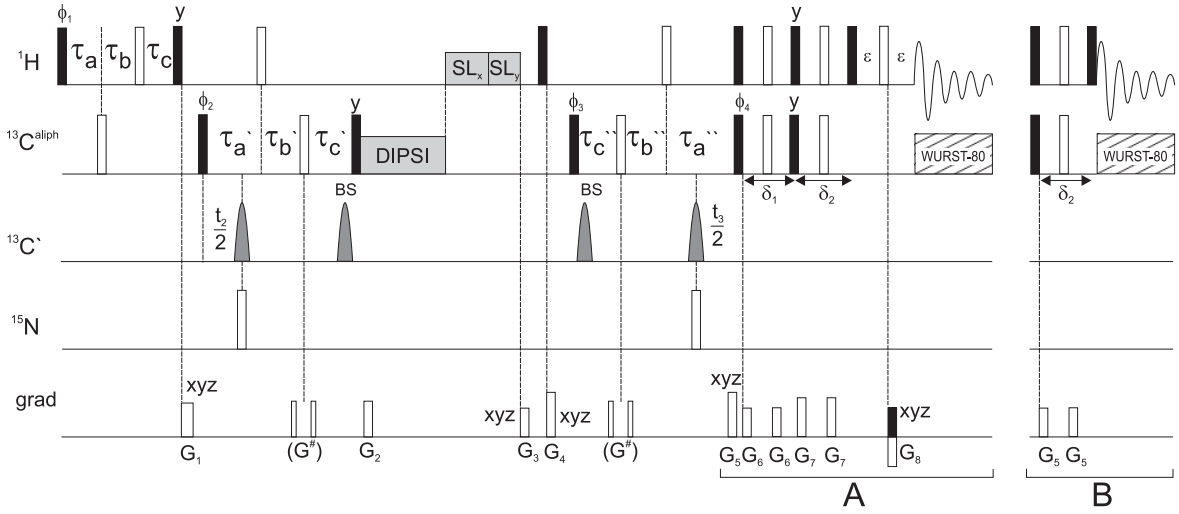


Fig. 4.1 Pulse scheme of 4D HCCH-TOCSY experiment. Sensitivity-enhanced version (A) could be applied if correlations to H^{α} are crucial, however, usually a standard $^{13}C \rightarrow {}^{1}H$ INEPT back-transfer is employed (B). The delays for shared-time evolution of $^{1}H(\omega_1)$ are as follows: $\tau_a = \frac{1}{2}(t_1 + \Delta)$, $\tau_b = \frac{1}{2}\, t_1\, (1 - \Delta/t_{1,max})$, $\tau_c = \frac{1}{2}\, \Delta\, (1 - t_1/t_{1,max})$, where $\Delta = (2\, J_{CH})^{-1}$ and $J_{CH} \approx 140$ Hz. The delays for SCT evolution in t_2 and t_3 (^{13}C) optimize the coherence transfer efficiency for CH_2 groups (0.71 for CH, and 0.92 for CH_3 groups, respectively): $\tau'_a = \frac{1}{2}(t_2 + \frac{1}{2}\Delta)$, $\tau'_b = \frac{1}{2}\, t_2\, (1 - \frac{1}{2}\Delta/t_{2,max})$, $\tau'_c = \frac{1}{4}\, \Delta\, (1 - t_2/t_{2,max})$, $\tau''_a = \frac{1}{2}(t_3 + \frac{1}{2}\Delta)$, $\tau''_b = \frac{1}{2}\, t_3\, (1 - \frac{1}{2}\Delta/t_{3,max})$, $\tau''_c = \frac{1}{4}\, \Delta\, (1 - t_3/t_{3,max})$. For the SE variant ($A$) $\delta_1 = 0.53\, \delta_2$ and $\delta_2 = (2\, J_{CH})^{-1}$ are set for a balanced coherence transfer for CH, CH_2 and CH_3 groups. Optional *gray* $^{13}C'$-selective inversion pulses refocus $J_{C\alpha C'}$-couplings and utilize Q3 profiles (Emsley and Bodenhausen 1992) with a phase modulation of −139 ppm with respect to the ^{13}C carrier freq. (35 ppm).

recognized by Coggins and Zhou (2008), who demonstrated it for relatively small GB1 protein (56 a. a.). Apart from using rather non-standard triple-sensitivity-enhanced version of the experiment, which is prone to sequence imperfections if limited phase-cycling is possible, Coggins and Zhou reported the efficiency of FFT-CLEAN algorithm in a rather imprecise way („average" 44%, and 78% at best, decrease of the apparent noise level, corresponding to reduction of 2.4 to 1.4 times the thermal noise level). What is more important, however, by sampling to t_{1max}=12.8 ms and $t_{2max} = t_{3max}$=6.1 ms Coggins and Zhou did not utilized the maximum allowed evolution space, possibly due to small resolution requirements posed by GB1 and primarly aiming at fast processing of the spectrum (with relatively small digital resolutions of 128 × 128 × 128 × 128 pts.).

The development of 4D SSAlgorithm encouraged us to present a far more challenging application of 4D HCCH-TOCSY (Stanek et al 2012). It was performed for partially disordered N-terminal fragment of Engrailed 2 protein (Augustyniak et al 2011), with NUS to 15 ms and 10 ms for indirect ^{1}H and ^{13}C evolutions, respectively. It should be mentioned that sampling of ^{13}C evolution is severely limited by up to three $^{1}J_{CC}$ couplings ($t_{\max} \leq (2^{1}J_{CC})^{-1}$). The performance of SSA appeared noticeably better than FFT-CLEAN, with an average reduction of noise level of 55% (in the ω_4 region between -1 and 6 ppm), and 90% at best. Since these average spans over both populated and empty spectral regions in ω_4, we suggested to focus attention on the representative region of intense resonances between 1 and 3 ppm. In this case the noise intensity was reduced by 71% on average (from 4.1 to 1.2 times the thermal noise). Even better artefact suppression was later obtained with improved SSAlgorithm (see sec. 3.4).

It is vital to investigate whether these improvements in noise suppression provide real

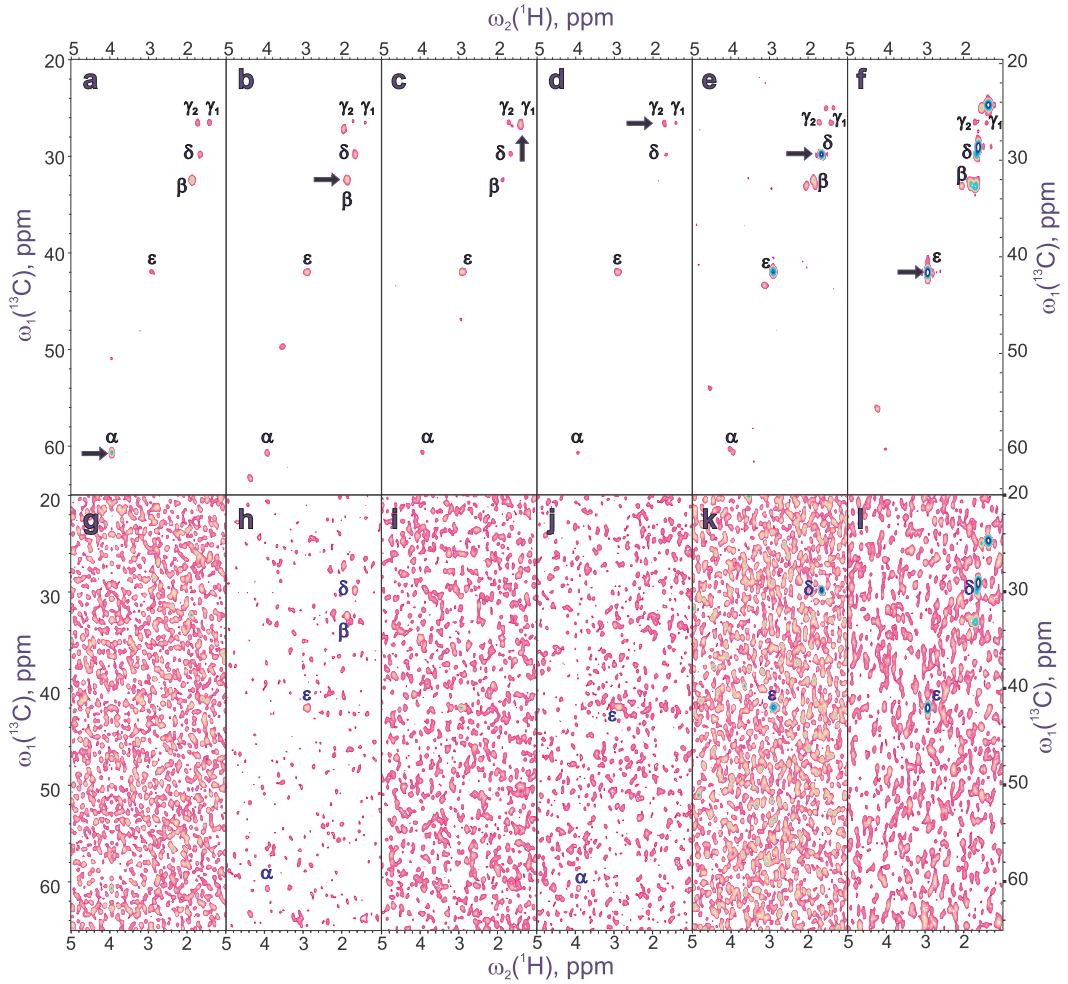


Fig. 4.2 A set of ^{1}H(ω_1) -^{13}C(ω_2) planes from 4D HCCH-TOCSY of Engrailed 2 homeoprotein, showing TOCSY cross-peaks to α (a, g), β (b, h), γ_1 (c, i), γ_2 (d, j), δ (e, k) and ϵ (f, l) ^{13}C-^{1}H(ω_3, ω_4) resonances of Lys216. The auto-correlation peaks ($\omega_1 = \omega_4$, $\omega_2 = \omega_3$) are indicated with arrows. The spectra obtained using SSA ($a - f$) and z.-a. FT ($g - l$) are plotted using identical intensity scale. Only resonances with a S/N $\geq$ 6 are labelled.

benefits in terms of peak detectability. For this purpose, a representative ω_1(^{1}H)-ω_2(^{13}C) planes for Lys216 of Engrailed are shown in Fig. 4.2. The comparison of spectra obtained with z.-a. FT and SSA demonstrates that while in the former none or very few peaks are above the noise level, in the latter nearly complete spin systems are usually present. Noteworthy, this enables to remove ambiguities (arising if two spin systems give correlations to neighbouring peaks) and utilize the data redundancy of 4D HCCH-TOCSY to identify cross-peaks which, for some independent reasons, are not present in all planes, e.g. for extended spin systems of Ile, Leu, Lys, Pro etc.

The influence of NUS artefact suppression on S/N was quantified for a similar spectrum of ubiquitin. The complete list of side chain resonances was used to auto-generate 2100 (!) possible cross-peaks in the 4D HCCH-TOCSY spectrum of this relatively small protein. The histograms shown in Fig. 4.3 clearly demonstrate that, to a large extent, the experiment is *sampling limited*. Nearly 1/4 of resonances are below a critical S/N, and the distribution is fairly sharp, with most resonances not exceeding S/N of 50 for z.-a. FT spectrum. On the contrary, the S/N is much diversified, with a substantial number of exceptionally strong

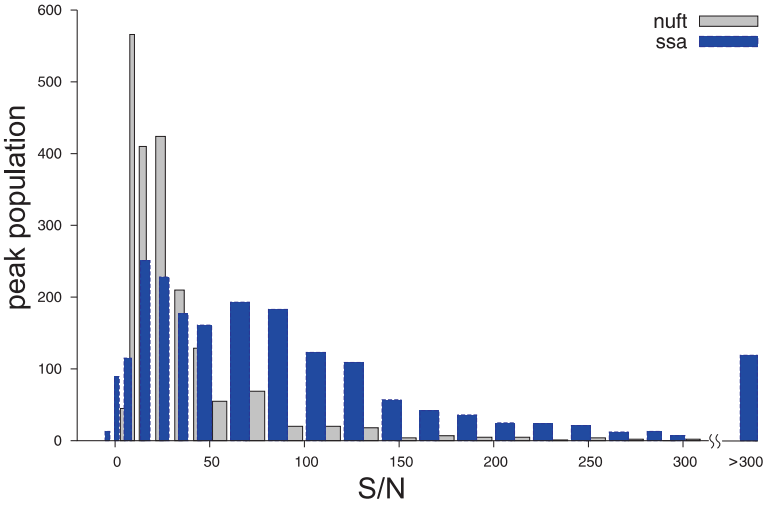


Fig. 4.3 The histograms of S/N for 2103 peaks in 4D HCCH-TOCSY spectra of human ubiquitin, obtained from the same data employing z.-a. FT (*gray*) and SSA (*blue bars*). The peak list was automatically generated from fully assigned peak list of ^{13}C-HSQC spectrum, giving all possible correlation peaks across a side-chain spin system, regardless of real TOCSY transfer functions.

resonances (S/N > 300) in the case of SSA processed spectrum. The detrimental effect of NUS artefacts in HCCH-TOCSY can be attributed to accumulation of artefacts stemming from a number of resonances sharing similar ω_4. Although most resonances have reasonable S/N $\geq$ 50 after a few hours of acquisition, the contributions of artefacts from m signals increase noise level by a factor of $\sqrt{m}$, usually obscuring all but the strongest peaks.

So far, the described approach was successfully applied for numerous protein samples, including Engrailed 2 (Dr. R. Augustyniak, formerly ENS Paris), a membrane protein YM2-MC (provided by Prof. O. Zerbe, Zürich University), two mutants of a homodimeric S100A1 protein, 2×93 a. a. (Dr. M. Nowakowski (2011) and Dr. K. Ruszczyńska-Bartnik, IBB PAS Warsaw), 169 a. a. EMAP-II and 181 a.a. LipazeA proteins. In two latter cases 4D HCCH-TOCSY played a crucial role in completion of side-chain assignment (Lozhko et al 2013; Augustyniak et al 2013).

The advantage of HCCH-TOCSY is that is allows to identify spin systems of residues which precede (otherwise inaccessible) a. a. with H^N exposed to chemical exchange. Noteworthy is also that HCCH-TOCSY can be optimized for significantly larger proteins employing MQ coherences (Zheng et al 2004) or ^{1}H-decoupling for SQ methyl ^{13}C evolution periods (Yang et al 2004). The approach explores superior relaxation properties of methyl resonances even in large proteins. In conjunction with 4D C,N-edited NOESY, which also preserves reasonable sensitivity for high molecular mass systems, it allowed to establish a general strategy for side-chain assignment without deuteration (Xu et al 2005, 2006). These methodological advancements emphasise the role of NUS for resolution enhancement of 4D HCCH-TOCSY, which could prove particularly beneficial for complex proteins. The application of SSA seems crucial in this case, otherwise any benefit from attenuated relaxation is offset by NUS artefacts.

4.1.2. 4D HCCH-TOCSY (ribose)

^{13}C-^{13}C TOCSY transfer is also commonly used for correlating ^{1}H and ^{13}C resonances in ribose rings in RNAs (Fürtig et al 2003). The pulse sequence was only slightly modified w. r. t. protein version (see Fig. 4.4), and applied to 34-nt RNA construct (named *LCS1co*) provided by Dr. M. Cevec from Slovenian NMR Centre (Cevec et al 2010). As argued in section 1.2.1, ribose resonances exhibit very limited chemical shift dispersion. Thus, high resolution in the key region of H1'/C1' region is of particular relevance. Similar restrictions associated with one or two $^1J_{CC} \approx 35$ Hz apply to resolution of ^{13}C dimensions. Within these

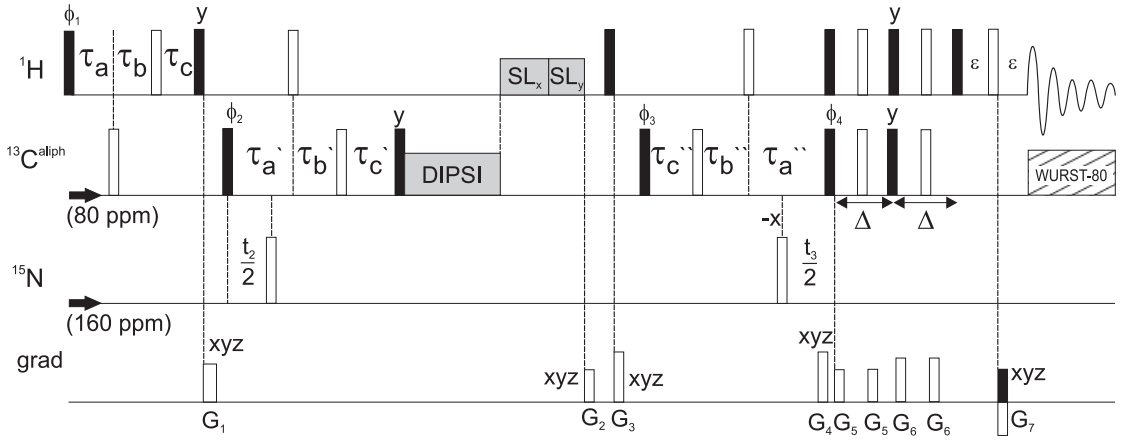


Fig. 4.4 Pulse sequence scheme of 4D HCCH-TOCSY for (deoxy)ribose ^{13}C-^{13}C correlations in RNA (DNA). Except for very big RNAs, sensitivity-enhanced scheme is preferred for evolution in t_3 (^{13}C), since all moieties (except H5"/H5'/C5') are CH groups, i. e. of multiplicity 1. $^1J_{\mathrm{CH}}$=150 Hz is assumed for coherence transfer delay $\Delta = (2J_{CH})^{-1}$. Timing of t_1 (^{1}H) evolution is as in the sequence in Fig. 4.1. Anti-phase to in-phase $\mathrm{H}_z\mathrm{C}_\pm \rightarrow \mathrm{C}_\pm$ coh. transfer delay in t_2 should remain a compromised value of $\frac{\Delta}{2}$ to preserve correlation *from* CH_2 groups (H5", H5'). In contrast, analogous delay embedded in t_3 evolution period could be optimized for CH groups (Δ) to enhance essential directly detected H1'/C1' region.

limitations, NUS allowed to obtain superior resolution of the 4D HCCH-TOCSY spectrum, since evolution was sampled up to 25 (^{1}H, t_1) and 10 ms (^{13}C, t_2 and t_3). In the moderate experimental time of 45 h only a small fraction of 0.3% of samples was collected, compared to conventional acquisition at these conditions (700 MHz ^{1}H frq., spectral windows of 3, and 11 kHz for ^{1}H and ^{13}C).

The statistics shown in Fig. 4.5 justify the application of SSAlgorithm in this case. The experiment seems *sampling limited* for RNA samples at typical experimental conditions, and suppression of NUS artefacts leads to roughly 2-fold average increase in sensitivity.

The simplicity of assignment process is illustrated in Fig. 4.6. If H1'/C1' peak is reasonably isolated a single plane from 4D spectrum suffices to assign all resonances in the ribose spin system. In the case of peak degeneracy, the assignment should be confirmed using additional planes with cross-peaks to H2'/C2', H3'/C3', etc. The comparison between z.-a. FT and SSA-processed spectra shows that NUS artefacts usually obscure less intense H5/H5"–H1' correlations. This entails quite inconvenient and time-consuming peak assignment process based on other available cross-sections. Thus, from a practical point of view, higher informa-

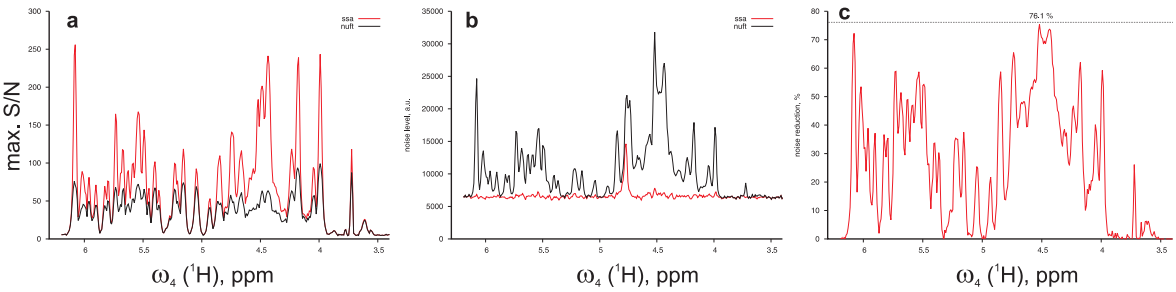


Fig. 4.5 Sensitivity (a) and noise level statistics (b, c) compared for 4D HCCH-TOCSY spectra processed with SSA (*red*) and z.-a. FT(*black curves*).

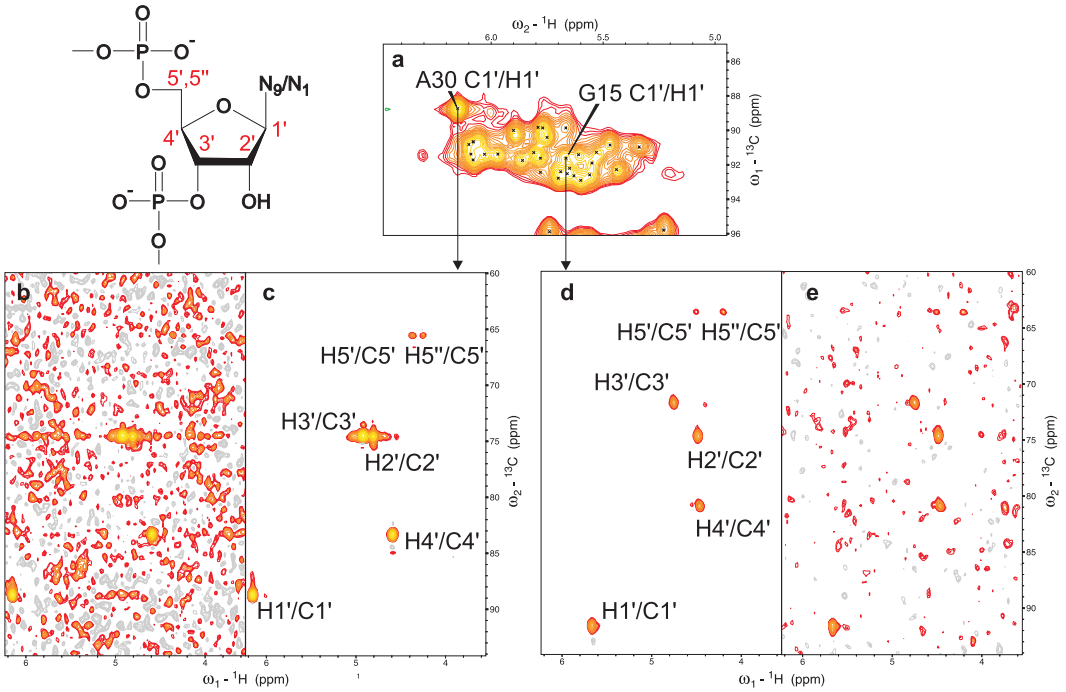


Fig. 4.6 $^{13}C(\omega_2)$–$^{1}H(\omega_1)$ planes from 4D HCCH-TOCSY spectrum for a 34-nt RNA, which show ribose resonances of A30 (a, c) and G15 (d, e) nucleotides. The starting point for analysis is 2D ribose H1' region of ^{13}C-HSQC (a). Compared are spectra obtained with SSA (c, d) and z.-a. FT (b, e). Identical intensity scale was used for all cross-sections. The noise level in (b) and (e) is elevated by a factor of 3.4 and 2.0 with respect to (c, d). Consequently, the weakest H5" & H5' resonances are missing unless NUS artefacts are removed.

tion content of SSA spectrum simplifies and increase efficiency of (usually manual) resonance assignment.

It is worth mentioning that even better resolution could be obtained by transferring coherence to nucleobase, and exploiting good dispersion of H6/H8, C6/C8, and possibly also N1/N9 chemical shifts. Due to increased coherence transfer delays, 4D/5D HC(CC-TOCSY)HCN would benefit from advanced relaxation optimizations, including MQ and TROSY ^{13}C-^{1}H coherences. This possibility will be investigated by a Ph. D. student in our group, Saurabh Saxena.

4.1.3. 4D HC(CC-TOCSY)(CO)NH

The ambiguities that may arise in 4D HCCH-TOCSY spectrum can usually be resolved by transferring coherence to backbone H^N for detection:

$$^{1}\mathrm{H}\ \ (\mathrm{SCT}\ t_1) \xrightarrow{^{1}J_{\mathrm{CH}}}\ ^{13}\mathrm{C}(\mathrm{SCT}\ t_2) \xrightarrow{\text{TOCSY mixing } (^{1}J_{\mathrm{CC}})}\ ^{13}\mathrm{C}^{\alpha} \xrightarrow{^{1}J_{\mathrm{C\alpha C'}}} {}^{13}\mathrm{C}' \xrightarrow{^{1}J_{\mathrm{C'N}}}\ ^{15}\mathrm{N}\ (\mathrm{SCT}\ t_3) \xrightarrow{^{1}J_{\mathrm{NH}}}\ ^{1}\mathrm{H}(t_4)$$

This scheme is particularly useful for stably folded proteins owing to usually sufficient peak dispersion in an ^{15}N-HSQC spectrum. This fact was recognized simultaneously by several groups, and 3D (Montelione et al 1992; Grzesiek et al 1993) and 4D correlations were proposed (Logan et al 1992; Clowes et al 1993). The undoubted simplicity of the assignment using 4D version is, however, obscured by the fact that *uniform* sampling leads to either

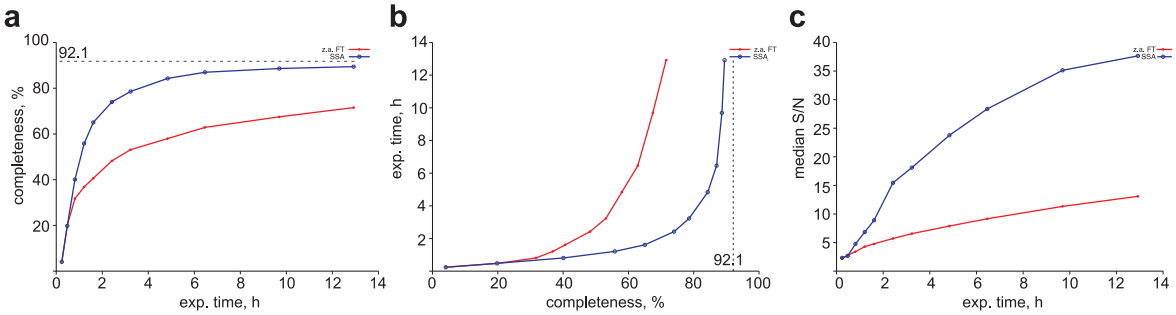


Fig. 4.7 Progress of side chain resonance assignment of human ubiquitin during 4D HC(CC-TOCSY)(CO)NH experiment. (a) Percentage of peaks above critical S/N=6 vs. experimental time in z.-a. FT (*red*) and SSA processed spectra (*blue curve*). Panel (b) shows a transposition of (a). 29 resonances (7.9 %) are inaccessible from this experiment because of chemical exchange and absence of proline H^N. (c) Increase of S/N of peaks during the experiment. The median of S/N ratio of all peaks is shown.

impractically long experimental times *or* poor resolution. Conventionally, the solution proposed by Grzesiek et al is used: a pair of high resolution 3D spectra is recorded, which include either ^{1}H or ^{13}C evolution. Nevertheless, for several a. a. the correlation of ^{1}H and ^{13}C chemical shifts is ambiguous. This results in frequent errors in assignments of peripheral side chain spins. The issue has been attracting considerable attention, and different sparse sampling approaches were employed, including projection-reconstruction (Jiang et al 2005), co-evolution (reduced-dimensionality) with 3D random sampling (Sun et al 2005), and finally, 4D NUS and MaxEnt processing (Mobli et al 2010).

The impact of NUS artefacts in randomly sampled 4D HC(CC-TOCSY)(CO)NH was investigated on an overnight experiment for a sample of human ubiquitin. In contrast to most examples in this chapter, we focus here only on the ultimate aim of the experiment: the number of detectable cross-peaks, or, equivalently, the completeness of assignment. The collected data set was divided to chunks which correspond to an increasing experimental time, and processed using z.-a. FT and SSA. Next, a complete list of 369 resonances at their theoretical positions (derived from this and other experiments) were loaded into each spectrum. Finally, S/N ratios were estimated for all peaks taking into account the variable noise level in different cross-sections of 4D spectra. The cross-peaks were considered detectable if S/N $\geq$ 6. The gradual build-up of information is compared for z.-a. FT and SSA spectra in Fig. 4.7a. Both curves should resemble a branch of hyperbole because of $\sqrt{t}$ scaling of both S/A and S/N ratios. At the initial stage both methods perform similarly. However, if SSA is employed the *plateau* is reached considerably faster, and most resonances are present after $\approx$ 12 hours of acquisition. In contrast, z.-a. FT generates a quite incomplete peak lists for this experimental time, and further acquisition seems very ineffective and impractical. The distance between the curves can be attributed to information loss caused by NUS artefacts.

It is interesting to analyse the inverse relationships (Fig. 4.7b), i.e. the time necessary to obtain a certain level of assignment. The vertical distance between curves can be interpreted as experimental time savings. Comparable lists of resonances are obtained using SSA in a significantly shorter time, and the speed-up factor increases with an assumed (target) completeness level. Obviously, the maximum number of detectable resonances in 4D HC(CC-TOCSY)(CO)NH is somewhat limited by absence of H^Nin prolines and fast exchange of some amide protons (6 a. a. residues for ubiquitin, including terminal Gly76).

Noteworthy is also the time dependence of peak S/N shown in Fig. 4.7c. Median of S/N

ratios was chosen since is effectively discards outliers and thus better reflects the overall peak population. While in the case of z.-a. FT spectra the S/N follows the $\sqrt{t}$ scaling, for SSA it increases faster, especially at the early stage of acquisition. Since signal accumulation is simply linear for both z.-a. FT and SSA, the unusual behaviour of S/N can only be attributed to NUS artefacts. At the beginning of experiment very few peaks can be identified by SSAlgorithm, and NUS artefacts are not suppressed effectively. As most of resonances start to exceed threshold S/N (say 10), the influence of artefacts is diminished, leading even to a decrease of effective noise (instead of an usual $\sqrt{t}$ accumulation). Once most peaks are above critical S/N, effective noise in SSA spectrum increases with time as thermal noise, consequently, S/N again has a $\sqrt{t}$ dependence.

To sum up the above discussion, the deleterious impact of NUS artefacts is also present in 4D HC(CC-TOCSY)(CO)NH experiment. The utility of SSA in this case lies in the ability to obtain the same results in a considerably shorter time. One should emphasise that the experiment is a method of choice for side chain assignment in small and middle size proteins. Depending on the intrinsic sensitivity the role of NUS artefacts may vary, however, even moderate time savings are valuable, particularly for samples of limited stability. Yet another practical aspect should be mentioned here: the SSA processed spectra feature fairly uniform noise level across the entire relevant spectral region. Thus, it is suitable for automated peak picking routines incorporated in most visualization software. The tedious manual peak picking can be avoided to a large extent which translates into a considerable speed-up of the entire study of a protein.

We wish to mention that the presented approach was applied to partially disordered Engrailed protein within access to NMR infrastructure, a part of East-NMR project, and virtually complete side chain assignment was accomplished (Augustyniak et al 2011).

Due to quite long coherence transfer delays ($\approx$ 7 ms for $C^{\alpha} \rightarrow C^{\alpha}{}_{z}C'_{z}$, $\approx$ 28 ms for $C'_{z} \rightarrow C'_{z}N_{z}$ and $\approx$ 28 ms for $C'_{z}N_{z} \rightarrow N_{z}H_{z}$) the primary application of the experiment are small or moderately sized proteins. The resolving power of this experiment cannot be exploited for well folded proteins $> \approx$ 20 kDa since it does not yield satisfactory sensitivity. Clearly, in this case the experiment is rather *sensitivity-* than *sampling-limited*, and the use of SSA is pointless.

4.1.4. 5D HC(CC-TOCSY)CONH

5D HC(CC-TOCSY)CONH experiment (Hiller et al 2008b), which correlates aliphatic resonances with backbone C', ^{15}N and H^{N} chemical shifts, has proved an efficient technique for side-chain assignment particularly for IDPs (Orbán-Németh et al 2013; Geist et al 2013b; Kızılsavaş et al 2013; Motáčková et al 2010). The 5D spectrum inherits superior resolution of HNCO, allowing for separation of ^{1}H-^{13}C resonances of an a.a. spin system. However, at least two factors may hamper full assignment of an aliphatic side chain: (a) for a given choice of mixing time the TOCSY transfer function may vary considerably for $C^{i} \rightarrow C^{\alpha}$ transfers from individual ^{13}C nuclei, and (b) the intensity of all peaks might be severely affected by chemical exchange of detected H^{N} or relaxation that occurs during quite lenghty coherence transfer delays. Finally, the accumulation of sampling artefacts stemming from peaks resonating in the similar ω_5 (direct detection) region, which is particularly severe for IDPs, may render weaker cross-peaks undetectable.

The 5D HC(CC-TOCSY)CONH experiment was performed for a sample of GAP43 protein on a 800 MHz spectrometer with a cryoprobe. The acquired data were processed using SSA and z.a. FT (*reference method*), and results evaluated using statistical tools as previously. The effective noise level (Fig. 4.8b) varies greatly across H^{N} dimension suggesting

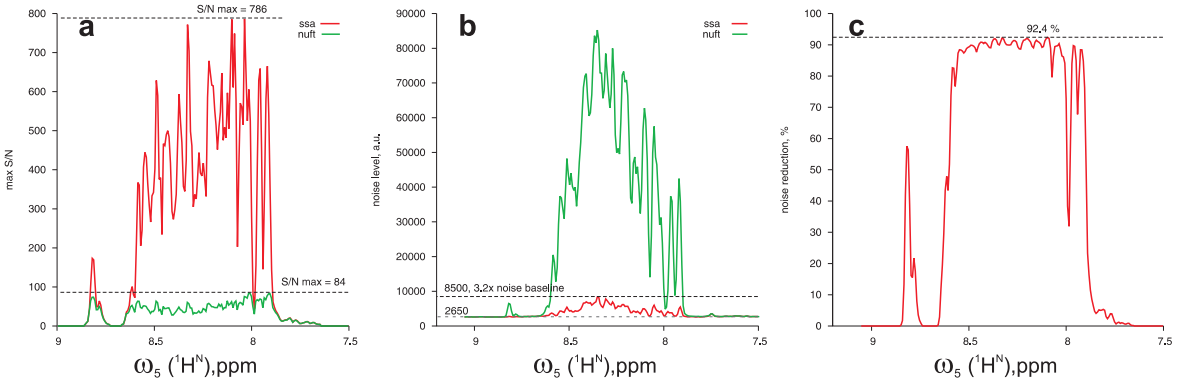


Fig. 4.8 Statistics of sensitivity (a), effective noise level (b) and its reduction by SSA (c) for the 5D HC(CC-TOCSY)CONH spectrum of GAP43.

that the experiment is mostly *sampling limited*, and majority of intrinsic sensitivity (Fig. 4.8a) can be recovered with the suppression of sampling artefacts. Noteworthy, despite the noise reduction is quite uniform for relevant ω_5 region and approaches 92.4%, the remaining contribution of residual artefacts is non-negligible (up to 220%). Nevertheless, obtained sensitivity enhancement corresponds to approx. 13^2=169-fold longer acquisition (19 h $\times$ 169 $\approx$ 133 days), which would be practically impossible, e.g. because of limited sample stability.

The significant improvement of information content of 5D HC(CC-TOCSY)CONH spectrum is demonstrated using $\omega_1(^1H^{ali})$-$\omega_5(^1H_N)$ and $\omega_2(^{13}C^{ali})$-$\omega_5(^1H_N)$ projections (Fig. 4.9), and two representative $\omega_1(^1H^{ali})$-$\omega_2(^{13}C^{ali})$ cross-sections for Arg7 and Ile25 (Fig. 4.10). The projections illustrate suppression of sampling noise globally, i.e. in the entire spectrum. Despite unavoidable simplification, Fig. 4.9 clearly shows that a considerable number of cross-peaks is buried in NUS artefacts unless they are removed using SSA. The more detailed inspection of selected cross-sections of Arg7 and Ile25 confirms this conclusion. Less intense resonances, such as Ile25 δ_1 or γ_1, and, in severe cases all cross-peaks (the case of

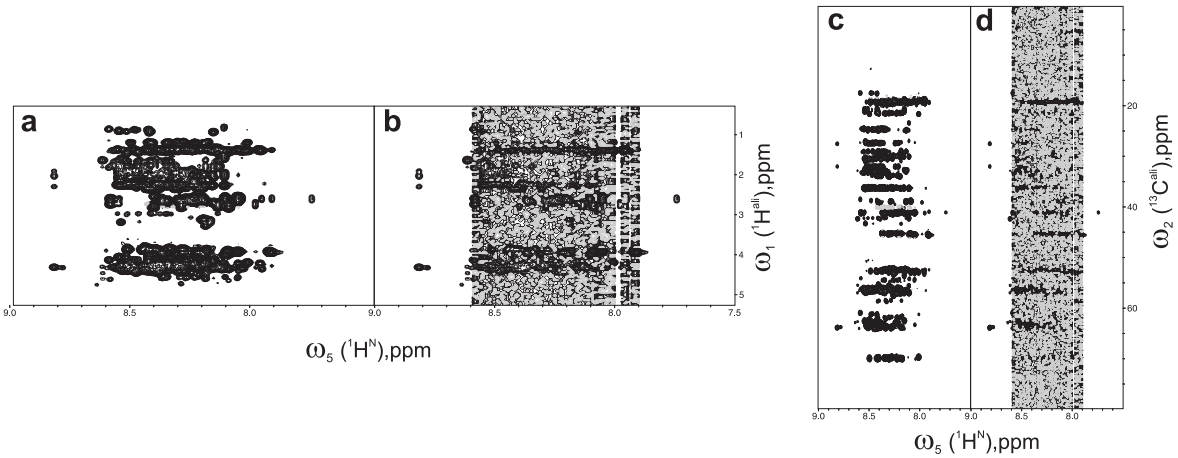


Fig. 4.9 Projections $\omega_1(^1H^{ali})$-$\omega_5(^1H_N)$ (a, b) and $\omega_2(^{13}C^{ali})$-$\omega_5(^1H_N)$ (c, d) from 5D HC(CC-TOCSY)CONH spectrum of GAP43, obtained with z.a. FT (b, d) and SSA (a, c) and plotted using identical intensity scale. These projections were selected (out of 4 indirect dimensions) since they show correlation peaks to various spins in the aliphatic side-chain. The comparison shows that weak TOCSY cross-peaks cannot be identified in the z.a. FT spectrum, and application of SSA leads to significant gain of information.

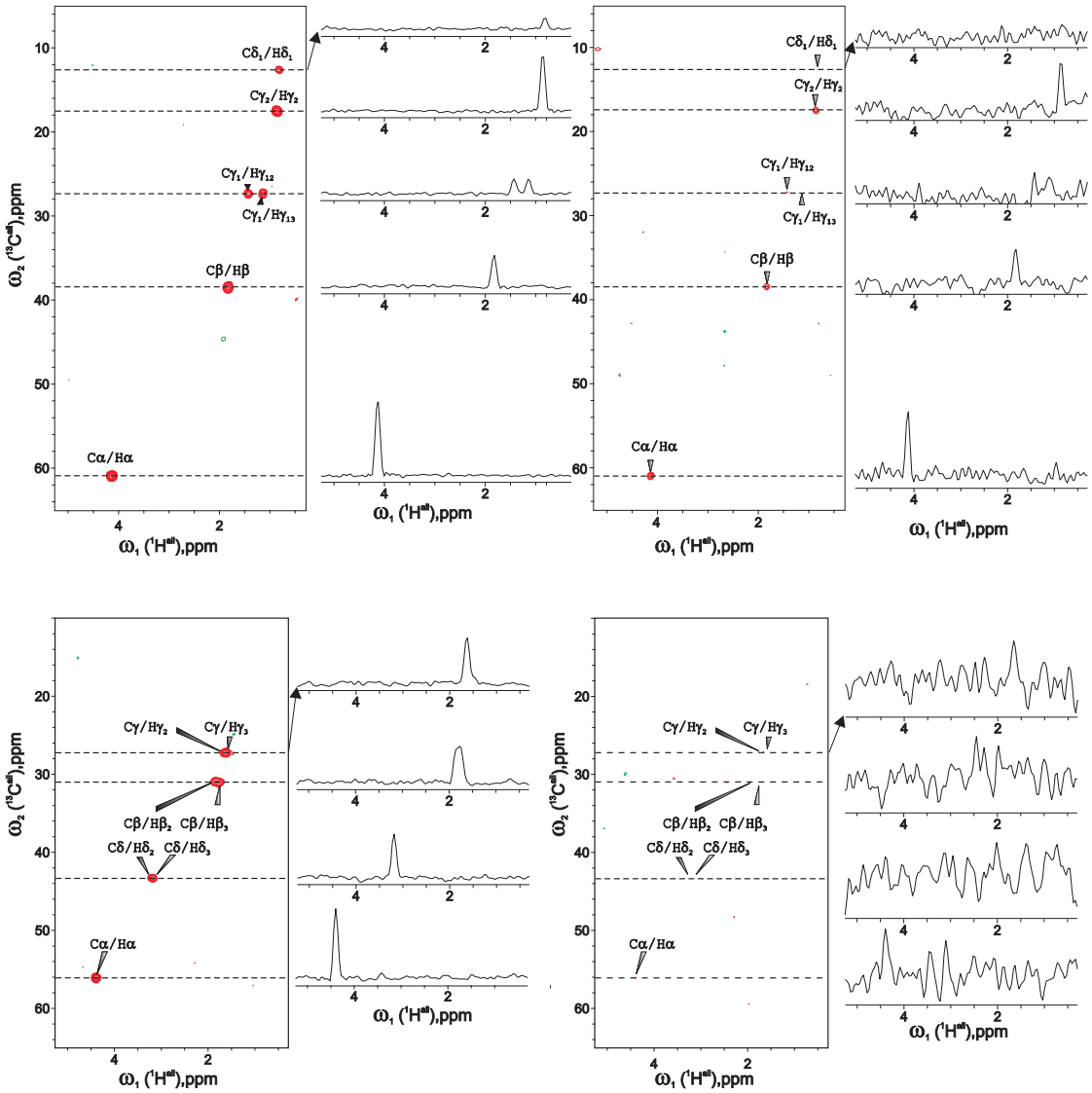


Fig. 4.10 $\omega_1(^1H^{ali})$-$\omega_2(^{13}C^{ali})$ cross-sections from 5D HC(CC-TOCSY)CONH spectrum of GAP43 showing resonances belonging to side chain spin systems of Ile25 (*top*) and Arg7 (*bottom panel*), obtained with SSA (*left*) and z.a. FT (*right panels*). The spectra are plotted using identical *relative* intensity scale (threshold of $4\times\sigma_{\mathrm{eff}}$). 1D (^{1}H) cross-sections are provided for all resonances, preserving intensity scale, to demonstrate S/N of peaks in the corresponding spectra. In the z.a. FT spectrum only the most intense resonances are detectable, however, ≈11.7-fold (Arg7) and 6.5-fold (Ile25) reduction of the effective noise level allows to identify all cross-peaks in the SSA processed spectrum.

Arg7), are below detection limit in the z. a. FT-processed spectrum but evident in the SSA counterpart. It is noteworthy, that, in contrast to 4D HCCH-TOCSY, there is no redundancy in the discussed spectrum. Consequently, it is of particular relevance to extract as many information as possible, and fully utilize the relatively high intrinsic sensitivity[1] of 5D HC(CC-TOCSY)CONH experiment. It should be empasised that – up to now – the capability of removing NUS artefacts in 5D spectra is *unique* for SSAlgorithm. In view of the promising applications such as shown above one anticipates it should be a valuable tool, and efforts will be made to publish it in the near future.

[1] compared to transfer experiments for sequential backbone resonance assignment

4.2. NOESY spectra for three-dimensional structure determination

4.2.1. Resolution enhancement of NOESY spectra

Nowadays, the NOE spectroscopy (NOESY) is the primary source of constraints for three-dimensional structure determination (Wüthrich 1986). It is essential to recognize that NOE scales with the interatomic distance as r^{-6} (see Eq. 1.13). This limits the available structural information to rather short-distance through-space contacts. Commonly, it is assumed that the presence of NOE cross-peak imposes an upper distance limit of about 5 Å. It is quite unlikely, or would require impractical acquisition time, to obtain NOE structural constraints beyond this limit. However, this fact should not necessarily be considered as a drawback.

The sphere of 5 Å radius contains a rather moderate number of other ^{1}H spins (say k on average). For a biomolecule containing n protons the number of cross-peaks in NOESY spectrum (N) is thus reasonable, and scales as $k \times n$, i.e. linearly with a system size. If n^2 all-to-all contacts would manifest in NOESY spectrum it would be hardly interpretable even for moderately sized biomolecules.

Nevertheless, homonuclear NOESY spectra may not suffice for structure determination even of small proteins due to ambiguity of cross-peak assignment. To illustrate the difficulty a simple model of this process is employed (Mumenthaler et al 1997). It is assumed that n ^{1}H resonances are evenly distributed over a frequency range of $\Delta\Omega$, and accuracy of peak position, $\Delta\omega$, closely related to ^{1}H linewidths in 2D spectrum, is common for all peaks. The probability that a resonance appears in the interval of $2\Delta\omega$ is thus $p = 2\Delta\omega/\Delta\Omega$. For the unique assignment of a peak in 1D ^{1}H spectrum only one resonance should match the given position, or, equivalently, remaining $n-1$ protons should not match. Its probability amounts to

$$p_{\text{uniq}} = (1-p)^{n-1} \tag{4.1}$$

To unambiguously assign a cross-peak this condition must be fulfilled for both ^{1}H frequencies, and the final probability is

$$p_{\text{uniq}} = (1-p)^{2(n-1)} \approx \exp(-2np) = \exp(-4n\Delta\omega/\Delta\Omega) \tag{4.2}$$

Let us assume that a typical number of well-isolated cross-peaks is N=$2\cdot10^3$ for a biomolecule containing $n = 500$ ^{1}H spins in a range of $\Delta\Omega = 10$ ppm, and linewidth is $\Delta\omega = 0.025$ ppm[2]. In this case the fraction of peaks that can be uniquely assigned is only 0.7%, corresponding to approximately 14 peaks. Obviously, this is insufficient data to start any automated structure calculation procedure without a reasonable preliminary structure. The ^{1}H resolution could be improved by extended sampling in the indirect dimension (say, up to $3T_2^*$), which is certainly at the cost of sensitivity, since the optimum for S/N is for $t_{\max} \approx 1.26\,T_2^*$ (Rovnyak et al 2004b). However, even under these conditions the natural linewidth limit is $\Delta\omega = (\pi\,T_2^*)^{-1} = R_2^*/\pi$. As delineated in Section 1.2.2, the relaxation rate scales linearly with a system size, $R_2^* \propto n$. Consequently, the exponential term in Eq. 4.2 decays at best as $\exp(-n^2)$ with increasing overall number of ^{1}H spins.

In conclusion, *de novo* structure determination of biomolecules requires a dramatic resolution enhancement. However, due to increasing ^{1}H linewidths (faster relaxation) the resolution can no longer be improved in homonuclear spectra. This has lead to heteronuclear-edited NOESY experiments, suitable for ^{13}C- and ^{15}N-labelled samples.

[2]For a small protein (say up to 100 residues) R_2^* (^{1}H) usually ranges up to 20 Hz allowing to obtain resolution of 20 Hz / 800 MHz = 0.025 ppm (at $B_0 = 18.8\ T$)

4.2.2. 3D ^{15}N- and ^{13}C-edited NOESY-HSQC

The single-heteronuclear-edited NOESY experiments were proposed by (Muhandiram et al 1993; Zhang et al 1994). They employ either HMQC or HSQC for ^{13}C-editing, and usually sensitivity-enhanced HSQC or TROSY for ^{15}N chemical shift evolution. The sensitivity cost is thus approximately $\exp(-2\tau_{HX}R_2^*(^1\mathrm{H}))$, where $\tau_{HX} = (2J_{HX})^{-1}$ is the time necessary to trigger coherence transfer while ^{1}H spins are in the transverse plane. For example, assuming that $\tau_\mathrm{C} = 3.7\ ms$, $\tau_\mathrm{N} = 5.4\ ms$, $R_2^*(^1\mathrm{H}) \approx 40\,\mathrm{Hz}$ the signal loss ranges to 26% and 35% for ^{13}C- and ^{15}N-editing. Additional cost of S/N per unit time associated with increased dimensionality was discussed in the Section 1.3.3, and futher losses are due to relaxation during evolution period of a heteronucleus.

Nevertheless, the reasonable signal dispersion is obtained owing to a limited correlation of ^{1}H and ^{15}N (or ^{13}C) chemical shifts. Hence, the assignment ambiguity is reduced significantly. Assuming that *destination* ^{1}H spin (i.e., the edited one) can be uniquely identified, the probablitily for unambigouos assignment of a cross-peak is governed by Eq. 4.1 rather than Eq. 4.2. This corresponds to a biomolecule of half the size ($n/2$) or a spectrum with twice the accuracy of ^{1}H chemical shifts (Güntert 2004).

However, conventional acquisition of 3D spectra is subject to sampling restrictions. In practice, a considerable amount of available resolution is sacrified to keep reasonable experimental time. E.g., NOESY-^{15}N-TROSY experiment requires approximately

$$N = N_1(^1\mathrm{H}) \times N_2(^{15}\mathrm{N}) = (t_{1,\max} \times 12\mathrm{ppm} \times \gamma_\mathrm{H} B_0)(t_{2,\max} \times 30\mathrm{ppm} \times \gamma_\mathrm{N} B_0)$$

samples. In order to keep it in the available experimental time, one has to decrease ^{1}H and ^{15}N evolution times (resolutions) as B_0 increases, thus losing resolution offered by modern high-field spectrometers. E.g., during 24h one could collect appox. 4300 increments, which allow for evolution up to barely 10 and 18 ms in ^{1}H and ^{15}N dimensions at $B_0 = 18.8$ T, respectively. Since these are far below respective T_2^* values, the *same* spectral resolution could be obtained at, say 500 MHz spectrometer as well. In essence, sampling limitations offset most benefits of both increased static field and dimensionality, as well as line-narrowing effect of TROSY.

NUS can be employed to overcome this difficulty. An overnight ^{15}N-edited NOESY experiment was performed for a sample of human ubiquitin, in which ^{1}H and ^{15}N were sampled up to 30 and 100 ms, corresponding to its relaxation properties. Compared to conventional acquisition the resolution was improved by a factor of ≈ 37. This is illustrated in Fig. 4.11, which shows the most crowded spectral region with cross-peaks for three overlapping amide protons (V26, E16, L71 $\mathrm{H_N}$). The major advantage in this case is the possiblity to isolate cross-peaks with enhanced precision of indirect ^{1}H chemical shift. Since peak picking is usually performed semi-automatically in 3D NOESY spectra, frequently misplaced or missing

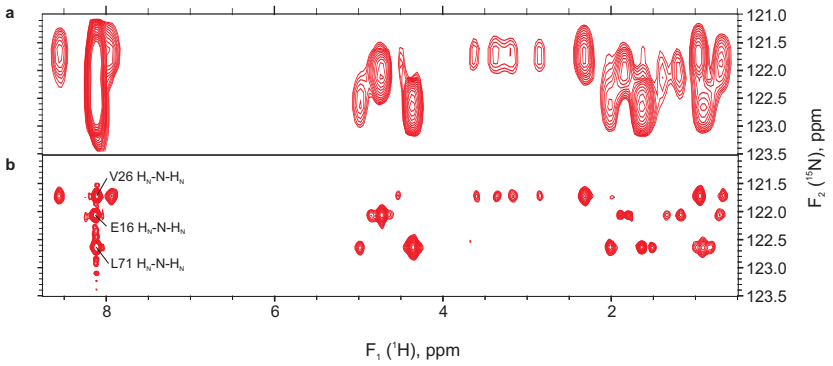


Fig. 4.11 Resolution enhancement in indirectly detected dimensions of 3D ^{15}N-edited NOESY of human ubiquitin. Compared are (a) conventionally and (b) non-uniformly sampled spectra *recorded in the same time*, processed with FT and SSA, respectively.

cross-peaks have to be found by a tedious inspection. For larger biomolecules, crowded regions are increasingly common, requiring more time and effort for peak picking. Moreover, signal overlaps inevitably result in decreased accuracy of ^{1}H peak positions, which is highly desired as argued before.

Efficient artefact suppression is crucial for NOESY spectra, since cross-peaks are at least two order of magnitude less intense than auto-correlation peaks. Effectively, a decrease of S/N ratio due to contribution of sampling artefacts could result in considerable loss of structural information. The ability of SSA to recover optimal sensitivity was investigated on the aforementioned ^{15}N-edited, and also ^{13}C-edited NOESY NUS data. In the latter case, ^{1}H and ^{13}C signals were indirectly sampled up to 30 and 12.5 ms with as few as 2700 points, yielding 15-fold resolution enhancement at the 16.4 T spectrometer. Figure 4.12 shows that SSA is capable of suppressing effective noise nearly to the thermal noise level. The noise reduction ranges to 94.4% and 98.4% for ^{15}N- and ^{13}C-edited NOESY spectra, respectively. The averaged values, which were computed for 70 and 369 diagonal peaks in these spectra, are 81.8% and 91.4%, respectively. These differences should be attributed to (a) higher sensitivity and (b) higher number of signals in ^{13}C-edited spectrum, which enhance the impact of NUS artefacts on the quality of spectrum.

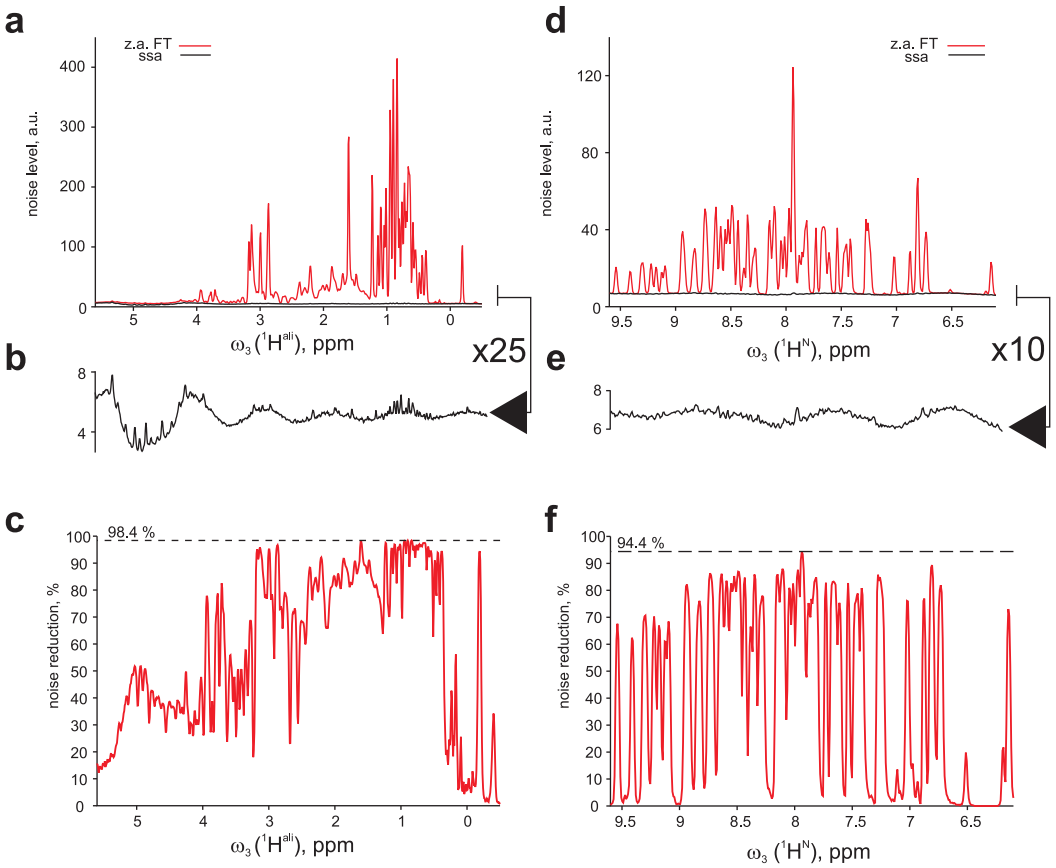


Fig. 4.12 Efficiency of NUS artefact suppression in 3D ^{13}C- (*a-c*) and ^{15}N-edited (*d-f*) NOESY-HSQC spectra of ubiquitin. *Top red* curves in panels (a, d) represent z. a. FT spectra while *bottom black* ones correspond to SSA. The latter curves are magnified in panels *b* and *e* to show that the influence of NUS artefacts in SSA spectrum is negligible with respect to, e.g. *wavy* distortions of baseline caused by imperfect digital solvent filter. The relative noise reduction is presented in panels *c* and *f*.

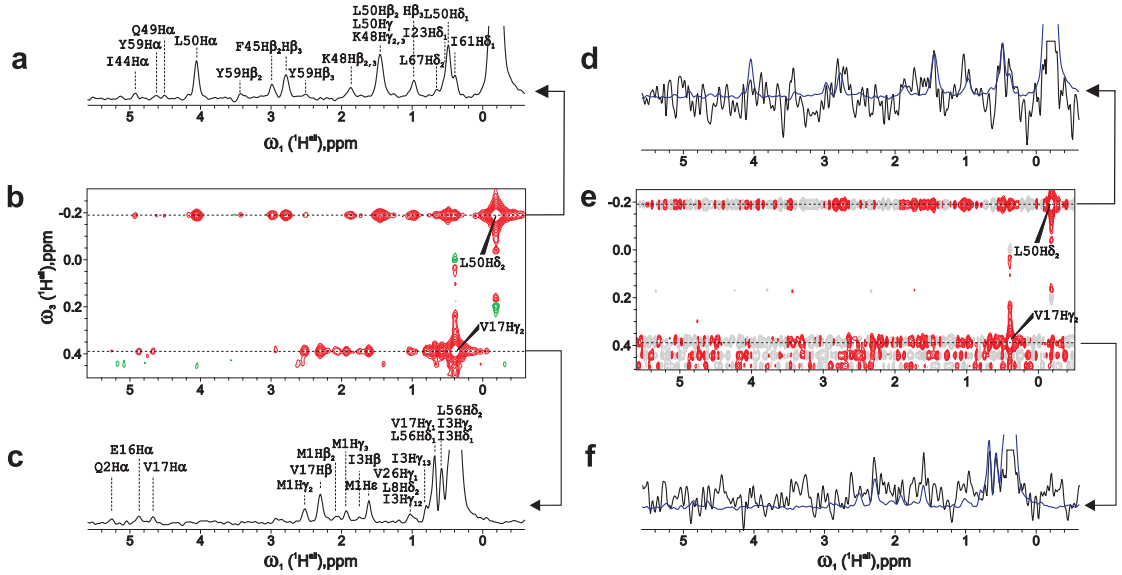


Fig. 4.13 The role of NUS artefacts in 3D NOESY spectra. SSA spectrum (a-c) is compared to z.-a. FT (d-f). Panels (b, e) show the ω_1-ω_3 (^{1}H-^{1}H) planes at ω_2=$\delta(^{13}\mathrm{C})$ = 19.6 ppm. Spectra are plotted using the same intensity scale, and the lowest contour is 0.2% of L50 Hδ_2 peak height. All observed cross-peaks are assigned and labelled on the 1D cross-sections (a, c). They are superimposed onto 1D cross-sections from z.a. FT spectrum (d, f) to enable comparison of NUS artefact level with the intensity of cross-peaks. Clearly, S/N of spectrum (e) is insufficient to recognize *any* NOE cross-peak.

The ^{13}C-edited NOESY spectra obtained with SSA and z.-a. FT are compared in Fig. 4.13. Two methlyl resonances, Val17 Hγ_2 and Leu50 Hδ_2, exhibit a number of relevant NOE cross-peaks (assigned) in SSA spectrum. On the contrary, *none* of them can be identified in the z.-a. FT spectrum as cross-peaks are completely obscured by NUS artefacts. These artefacts originate from only one (isolated) auto-correlation signal in both cases, and effective S/N is even lower in more crowded regions of the spectrum. This example shows that NUS NOESY spectra are utterly useless until artefacts are suppressed. It is noteworthy that despite considerably enhanced resolution in indirect ^{1}H (ω_1) dimension, some cross-peaks are still overlapped (Fig. 4.13a,c). To some extent, modern structure calculation protocols can cope with such moderately ambigouos restraints, which are deduced during consecutive iterations of concerted structure calculation and peak assignment algorithm.

To verify whether SSA reconstruction preserves genuine peaks without introducing *false positives*, two additional ^{15}N-edited NOESY spectra were recorded. The reference high resolution conventional spectrum was acquired in 7 days using 240 t_1 and 120 t_2 increments, yielding resolution of 50 Hz ($(20\ \mathrm{ms})^{-1}$) and 20 Hz ($(50\ \mathrm{ms})^{-1}$) in the indirect dimensions. The NUS data was collected assuming the same max. evolution times of 20 and 50 ms, and a fraction of θ=8% of samples used for conventional spectrum. Therefore the NUS spectrum had sensitivity lower by a factor of $\sqrt{\theta^{-1}}$=3.5. Since both spectra featured identical resolution it was possible to compare their information content. Threshold for peak detection was 10 and 35 of thermal noise level[3], σ, for NUS and US spectra, respectively. Different thresholds account for the difference in sensitivity between the NUS and the reference spectra. The peak was classified as properly reconstructed if its intensity in the reconstructed, I_{NUS}, and reference, $I_{\mathrm{ref}} \times \theta$, spectra differed by no more than $3\,\sigma_{\mathrm{NUS}}$. 828 (98.6%) peaks were correctly identified, and there were only 11 (1.3%) *false positives*. The reference spectrum showed

[3]determined as a median of 10^4 randomly chosen spectral intensities

14 additional peaks which should be considered as *false negatives* (missing peaks) of NUS spectrum. The linear correlation of peak intensities was demonstrated in Section 3.4. These results confirm that SSA generates correct NOESY spectra without affecting its structural information content (Stanek and Koźmiński 2010).

In conclusion, NUS can be successfully employed for resolution enhancement of challenging 3D NOESY spectra. Recent findings of Tikole et al (2013) suggest that, irrespectively of a molecular size, the accuracy of obtained three-dimensional structure was optimal if a natural linewidth in the indirect ^{1}H dimension was obtained using NUS. This quite surprising conclusion was drawn from a large scale survey among 400 deposited protein structures between 5 and 33 kDa. Despite these convincing results, NUS is rather rarely used in a routine setup of 3D NOESY spectra. One of the reasons could be that moderate resolution enhancement does not seem attractive enough to modify the setup of the experiment and employ more sophisticated data processing algorithms. Consequently, uniform sampling, even if known to be inoptimal in terms of obtained resolution, is chosen. Another reason for limited popularity of NUS in the context of NOESY is the need for careful data processing that should ensure (a) high dynamic range, and (b) linear signal response even in crowded spectral regions. Nowadays, several methods including MDD, *compressed sensing*, l_1-reconstruction, SSA and algorithms related to CLEAN have been proven useful to process such data.

4.2.3. 4D ^{13}C,^{13}C-edited NOESY

It has long been recognized that optimal utilization of ^{15}N,^{13}C-labelling for resolution enhancement is feasible by use of four-dimensional double heteronuclear-edited NOESY spectra[4] (Kay et al 1990a). As it was emphasized by Luan et al (2005), 4D NOESY spectra are a more natural solution and provide the best possible resolution, since both *source* and *destination* protons can reliably be identified on HSQC-type 2D correlation spectra. However, poor resolution that can be obtained using conventional sampling in a reasonable time offsets any benefit of higher dimensionality (Tugarinov et al 2005b), and moderate 3-4-fold aliasing of ^{13}C-dimension(s) (Bax et al 1990a; Kay et al 1989), often accompanied by *linear prediction* or *Maximum Entropy* extrapolation, can only slightly alleviate the difficulty.

The 4D ^{13}C,^{13}C-edited [^{1}H,^{1}H]-NOESY experiment was derived from the 3D couterparts by substitution of initial ^{1}H evolution period by HMQC (Clore et al 1991; Vuister et al 1993). HMQC is prefferred to HSQC because of a fewer number of ^{1}H and broadband ^{13}C inversion pulses, and also due to attenuated relaxation of MQ ^{1}H-^{13}C coherences (Norwood 1992; Grzesiek et al 1995). As shown later by Tugarinov et al (2003), HMQC also preserves slowly relaxing components in CH_3 moieties (also referred to as methyl-TROSY effect), thus making it suitable for large proteins. Recently, the improved indirect HMQC schemes for optimal utilization of *two* coherence transfer delays for shared-time ^{1}H evolution as described in Section 1.4.2 were proposed (Xu et al 2006; Stanek et al 2012). This optimization partially compensates the additional sensitivity cost of extending a pulse sequence, which is analogous to the transition from 2D to 3D single-edited NOESY discussed in Sec. 4.2.2.

A 4D ^{13}C,^{13}C-edited experiment was performed on a sample of human ubiquitin at very low relative sampling density of $\theta = 3 \cdot 10^{-3}$, which was possible due to no correlation of θ with the quality of reconstruction. The representative 2D $\omega_1(^1\mathrm{H})$-$\omega_2(^{13}\mathrm{C})$ cross-sections are compared to 1D $\omega_1(^1\mathrm{H})$ slices from 3D counterpart in Fig. 4.14. Clearly, the additional ^{13}C dimension allows to assign highly ambiguous cross-peaks, and also resolves overlapped ones into well separated peaks. Despite a relatively low complexity of ubiquitin (76 a.a.), some

[4]Interestingly, 4D C,N-edited [^{1}H,^{1}H]-NOESY was the first proposed four-dimensional experiment

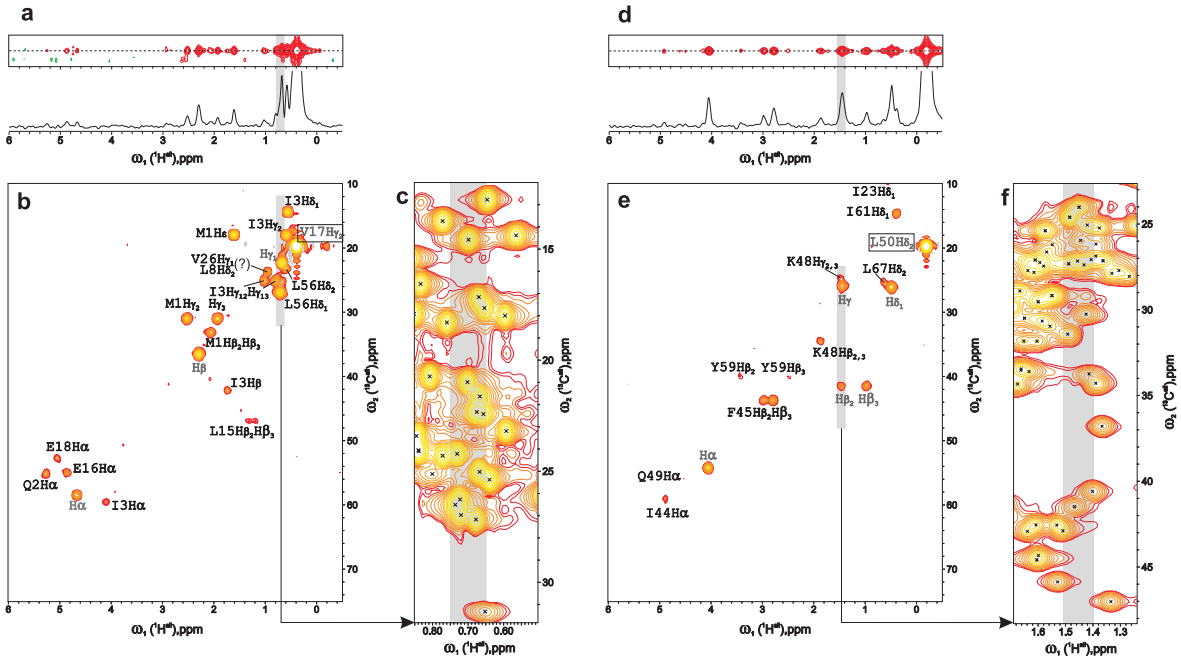


Fig. 4.14 The comparison of spectral resolution of 3D ^{13}C-edited and 4D ^{13}C,^{13}C-edited NOESY spectra of human ubiquitin. The 1D slices in panels *a,b* correspond to Val17 Hγ_2 and Leu50 Hδ_2 spins that were presented in Fig. 4.13. The visible correlations are spread in the additional ω_2(^{13}C) dimension in 4D spectrum (panels *b* and *e*). Particularly ambiguous and overlapped cross-peaks are *shaded*, and compared with corresponding regions of ^{13}C-HSQC (panels *c,f*). Within a relatively high precision of ^{1}H chem. shift of 0.05 ppm there are approx. 14 matching peaks in each case, and the ambiguity is completely removed in the 4D spectrum. A number of long- and middle-range (*black labels*) as well as intra-residue (*gray labels*) cross-peaks were unambiguously assigned owing to high resolution of 4D spectrum. The only ambiguous peak, denoted (?) in panel *b*, indicates Leu8 Hδ_2 or Val26 Hγ_1.

cross-peaks, enclosed in *gray* areas in Fig. 4.14, may have as many as 14 possible matching resonances as shown in the magnified HSQC regions in Fig. 4.14c,f. The 4D experiment dramatically improves the reliability of assignment since only 1 of the cross-peaks shown here has a multiple possible assignment (cf. Fig. 4.14b).

As for the 3D counterpart, an efficient suppression of NUS artefacts is essential. The SSA reconstruction is compared to z.a. FT spectrum in Fig. 4.15 on two orthogonal cross-sections which show correlations to Val17 Hγ_2 spin. This example demonstrates that virtually no relevant structural information could be obtained from 4D NUS data unless SSA is employed. A more comprehensive comparison is provided by 2D ^{1}H-^{1}H projections from 4D SSA and z.a. FT spectra (cf. Fig. 4.16). While SSA reconstruction is only slightly affected by residual NUS artefacts, in the z.a. FT counterpart almost all cross-peaks are completely obscured. This confirms that 4D ^{13}C,^{13}C-edited NOESY spectra are entirely (globally) *sampling-*, not *sensitivity-limited*.

The role of NUS artefacts and efficiency of SSA in their suppression was investigated in detail on three significantly different samples of well-folded proteins: a standard 1.0 mM sample of human ubiquitin (76 a.a., 8.5 kDa) purchased from ASLA, a 0.5 mM sample of a E32Q mutant of a homodimeric S100A1 protein (2×93 a.a., 20 kDa, PDB code 2LHL) and <0.5 mM sample of LipaseA (181 a.a, 19.1 kDa) from *Bacillus subtilis* (Augustyniak et al 2012, 2013). Two kinds of measure were employed: a percentage of effective noise reduction

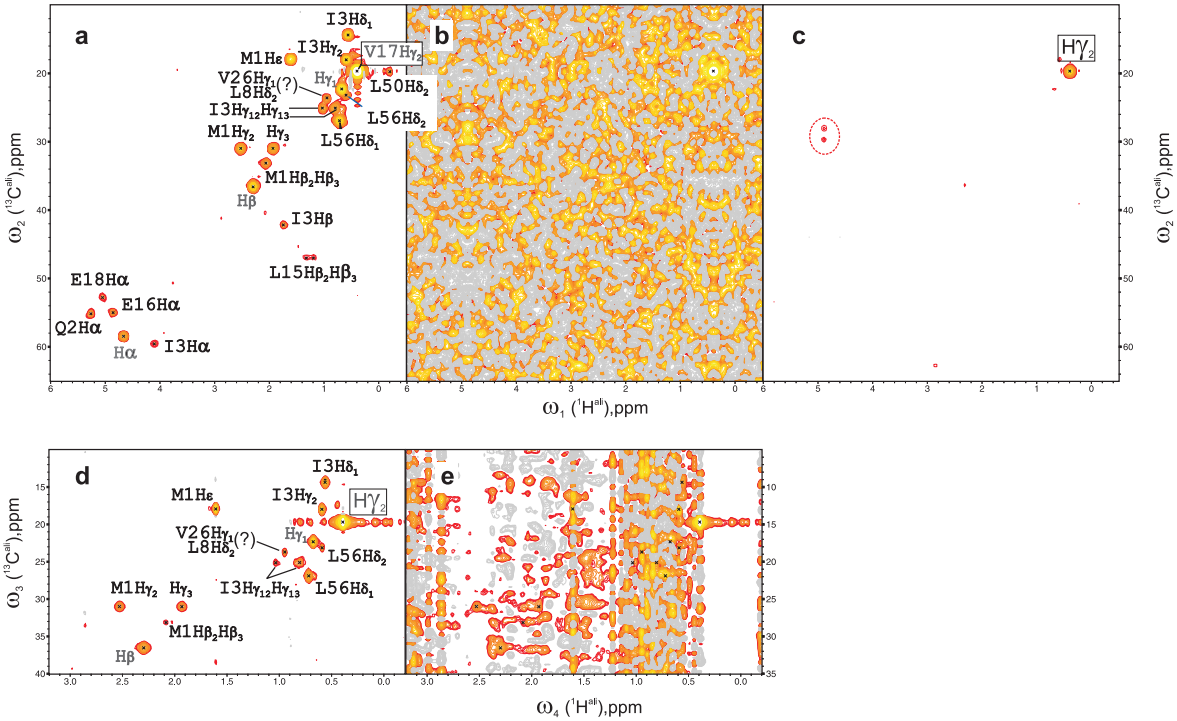


Fig. 4.15 ^{13}C-^{1}H cross-sections from 4D C,C-edited NOESY spectrum of ubiquitin (a-e), which show correlation peaks to Val17 Hγ_2. Panels (a-c) show the (*indirect*) cross-sections $\omega_1(^1\text{H})$-$\omega_2(^{13}\text{C})$ at (ω_3,ω_4) = Val17 Cγ_2, Hγ_2. Panels (d, e) illustrate *orthogonal* cross-sections, namely $\omega_3(^{13}\text{C})$-$\omega_4(^1\text{H})$ at (ω_2,ω_1) = Val17 Cγ_2, Hγ_2. Compared are the spectra obtained using SSA (a, d) and z.-a. FT (b, c, e); spectra (a) and (b), and (d) and (e) are plotted with pairwise identical intensity scale. The lowest contour in panels (a) and (c) corresponds to 4 × σ_{eff}, which is 34-fold larger for z.-a. FT spectrum. The cross-peaks are completely obscured by sampling artefacts in z.a. FT spectrum (b, c, e), but can be easily identified after artefact suppression (a, d).

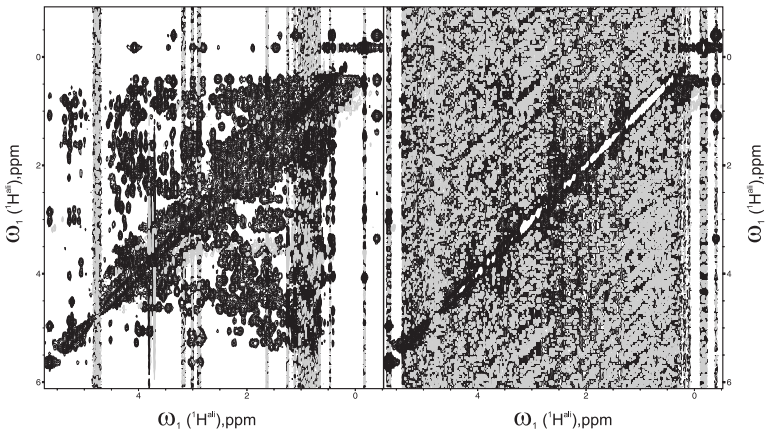


Fig. 4.16 $\omega_1(^1\text{H})$-$\omega_4(^1\text{H})$ projections from 4D C,C-edited NOESY spectrum of ubiquitin, processed with SSA (*left*) and z.-a. FT (*right panel*). Only few vertical stripes of residual artefacts are present in SSA spectrum, while almost no NOE cross-peaks can be identified in the reference one. Both projections are plotted using the same intensity scale (lowest contour at 10 σ_{thermal}).

(cf. Fig. 4.17a-c) and an effective noise level in z.a. FT and SSA spectra (Fig. 4.17d-f). In all three cases, the contribution of NUS artefacts is significant and ranges from 40 to 99% of effective noise. The performance of SSA is quite invariant, and observed variations in maximum noise reduction (between 98.2% for ubiquitin and 92.6% for LipA) are due to different relative contribution on NUS artefacts. The reconstructed spectra feature dramatically decreased noise level (cf. Fig. 4.17d-f). The intensity of NUS artefacts (in arbitrary units), compared to noise baseline, decreases in the order of UBQ > S100A1 > LipA. It should

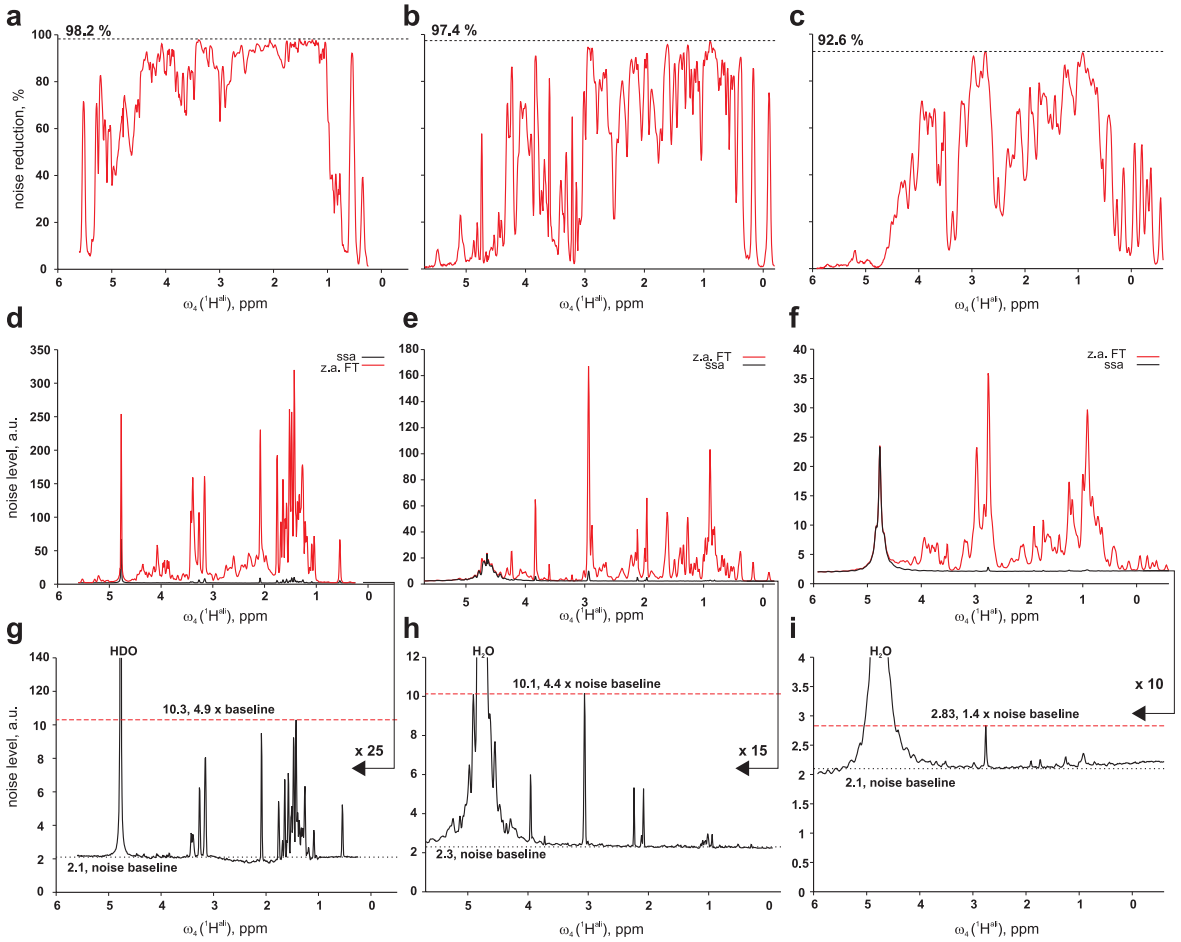


Fig. 4.17 Efficiency of artefact suppression in 4D C,C-edited NOESY spectra of ubiquitin (*a,d,g*), S100A1(E32Q) (*b,e,h*) and Lipase A(*c,f,i*), expressed in terms of noise level reduction (*a-c*), and noise level in z.a. FT (*d-f*, *red curves*) and SSA reconstructed spectra (*d-f*, also magnified in *g-i*, *black curves*).

be noted that this measure also reflects the relative intrinsic sensitivity of the experiments. The residual artefacts are still present after SSA reconstruction, and exhibit on magnified (note different factors) plots (Fig. 4.17g-i). In the case of ubiquitin, SSA reached its limits in suppression of artefacts and residual ones are still above thermal noise (up to $4.9\times\sigma_{\text{thermal}}$). In two remaining examples the influence of residual artefacts is smaller, suggesting that in typical applications, i. e. more diluted solutions and larger proteins, residual NUS artefacts do not play any significant role.

These results were confirmed by comprehensive statistics of noise and S/N ratios for *all* assigned side-chain resonances in three studied proteins. The results are summarised in Table 4.1. The positions of diagonal peaks, which also indicate location of relevant cross-peaks, were used to reflect non-uniform distribution of signals across direct ^{1}H dimension and, thus, the influence of NUS artefacts to average S/N. The application of SSA to 4D C,C-NOESY data yielded reduction of average noise by a factor of 19.3, 5.9 and 3.1 for UBQ, S100A1 and LipA, respectively. The noise is elevated by a factor of 1.2-1.4 with respect to a perfect reconstruction. Smaller contribution of NUS artefacts for LipA and S100A1 is a consequence of lower intrinsic sensitivity as shown in Table 4.2. Median S/N ratio for all diagonal peaks was normalized to equal acquisition time for a meaningful comparison. To

Table 4.1 Noise level in 4D NUS C,C-edited [^{1}H,^{1}H]-NOESY spectra.

Protein	N_{diag}[a]	Noise reduction		Average noise[b]		Max. noise[b]	
		average	max.	z.a. FT	SSA	z.a. FT	SSA
Ubiquitin	369	94.7%	98.2%	25.1	1.3	139.1	4.9
S100A1(E32Q)	401	83.3%	97.4%	8.2	1.4	71.6	3.9
Lipase A	725	67.4%	92.6%	3.7	1.2	17.1	1.4

[a]number of diagonal peaks which positions were used to compute statistics
[b]with respect to baseline noise level

isolate the influence of NUS artefacts the estimation of S/N in a perfectly reconstructed spectrum was also provided. The different sample conditions resulted in over 10-fold ratio of S/N between ubiquitin and Lipase A, which fully explains differences in average noise reduction (Table 4.1).

Protein	S/N median[a]		
	z.a. FT	SSA	perfect[b]
Ubiquitin	43	469	475
S100A1(E32Q)	50	167	194
Lipase A	12	34	36

Table 4.2 Average S/N of diagonal peaks in C,C-edited NUS spectra.

[a]Normalized to equal experimental times. S/N is proportional to $\sqrt{nt \times ni}$, where nt is the number of scans and ni is the number of indirect points. The shown data are normalized to $nt \times ni = 12400$.
[b]Assuming baseline noise level for all peaks

Subsequently, obtained 4D C,C-NOESY spectra were analysed regarding structural information content by Dr. W. Augustyniak (LipA) and Dr. M. Nowakowski and Dr. K. Ruszczyska-Bartnik (S100A1 E32Q). First, they were compared to conventionally sampled 3D spectra. Surprisingly, for some residues 4D spectra give more structural constraints than 3D counterparts since cross-peaks could be identified even if overlapped with autocorrelation signal in ^{1}H(ω_1) dimension (Fig. 4.18a,c) or clusters split into many individual cross-peaks (Fig. 4.18b,d). Additionally, observed cross-peaks were verified against previously determined three-dimensional structures (cf. Fig. 4.18 and also Fig. 4.28 on page 113). No systematic structure violations were detected. The data obtained for Lipase A from 4D C,C-edited NOESY spectrum were also used as input to structure calculation protocol, and results compared to the reference high quality structure determined previously using NMR. In general, the obtained fold was consistent with a reference, with backbone RMSD of mean structure to reference of 2.96 Å and 2.18 Å within a bundle of 20 conformers. The largest deviations were found in the extended loops and terminal regions, where, contrary to helices and strands, limited number of structural restraints were accessible. Although it was possible to derive a correct structural model from *standalone* 4D NOESY data, a joint use with lower-dimensionality spectra turned to be a more promising approach. In this case, ca. 2300 cross-peaks from the overnight 2D NOESY spectrum combined with as few as 185 manually assigned unambigouos distance restraints from 4D spectrum yielded a substantially more precise structure ensamble: backbone RMSD of 1.32 Å within a bundle and 1.8 Å w.r.t. the reference. The superposition of 20 conformers and the reference structure is shown in Fig. 4.19. The detailed analysis of information content of 3D and 4D NOESY spectra *recorded in the same time* was also performed (Augustyniak 2012). 4D C,C-edited spectrum yielded only slightly lower number of cross-peaks (9%) than a 3D counterpart, however, the data appeared significantly more redundant. In effect, nearly 42% less restraints were obtained, with concomitant loss of structure precision and accuracy. This result was attributed to inherently

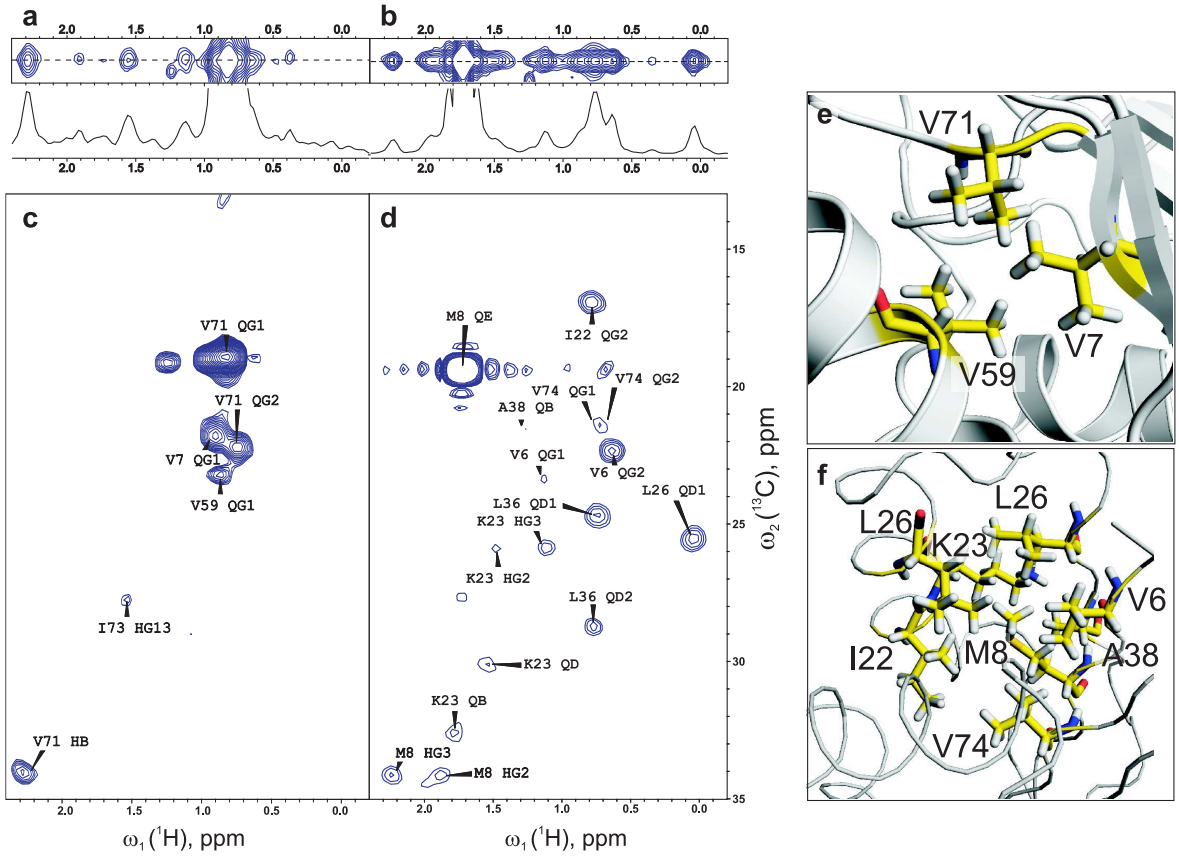


Fig. 4.18 Comparison of resolution between conventional 3D (*a, b*) and 4D NUS (*c, d*) ^{13}C-edited spectra of Lipase A. Observed cross-peaks are fully compatible with X-ray derived structure (*e, f*) (van Pouderoyen et al 2001). Because of lower sensitivity of 4D spectrum, some peaks present in 3D counterpart are below detection threshold. 3D structural models were prepared by Dr. W. Augustyniak.

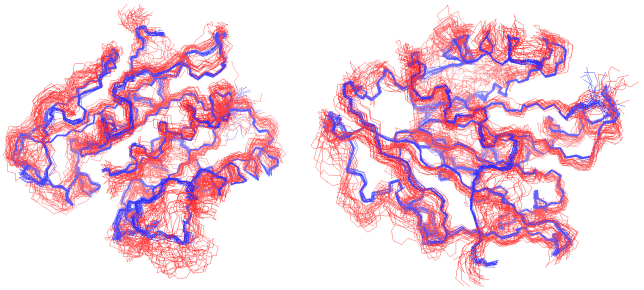


Fig. 4.19 Superimposed structures (20 conformers) of Lipase A determined using 4D NOESY spectra and supported by 2D experiment (*red*), and from a reference high-resolution model (*blue*). The latter one was obtained using significantly better hardware and a different set of structural constraints, thus the precision of structures cannot be directly compared. The figure courtesy of Dr. W. Augustyniak.

lower sensitivity of 4D experiments. It is also noteworthy that CYANA tends to derive more restraints from ambigouos 3D NOESY peaks than from 4D ones. In conclusion, for middle-size proteins such as Lipase A, the structure determination using 4D NOESY spectra turned to be an inoptimal solution. However, it proved useful at the initial stage of structure calculation, which required manual assignment of available unambigouos long-range restraints for a tentative model. 4D spectra provided 185 peaks, while 3D counterparts as few as 25 peaks. As discussed in Sec. 4.2.1, this figure is crucial for enforcing structure refinement towards a correct structural model. Indeed, from 3D data the automated protocol did not succeed, and manual intervention (peak assignment) was necessary in this case. The possible applications

of 4D NOESY for **middle size** proteins are thus:

- validation of strucures obtained with more ambigouos data
- determination of a reliable starting model, if impossible from 3D data
- automated high-troughput structure determination with minimal human intervention (see also recent work of Coggins et al 2012)

An interesing, yet not studied issue is whether 4D NUS NOESY data are more immune to incomplete side-chain assignments, which is a bottleneck of conventional approach.

4D NUS NOESY methods were shown essential for structure determination of **large proteins** in solution. The most spectacular examples include [ILV-^{1}H,^{13}C, otherwise ^{2}H]-labelled, H^{N}-back exchanged proteins with molecular mass ranging to 82 kDa (Malathe Synthetaze, 723 a.a.) (Tugarinov et al 2005b; Hiller et al 2009). For these systems a sensitive 4D C-HMQC-NOESY-C-HMQC, which exploits methyl-TROSY effect, was employed. NUS data were processed with MDD (Tugarinov et al 2005a), or co-processed with 2D HMQC spectra (Hiller et al 2009). Interesing discussion on the use of 4D data for large systems, which goes beyond of the scope of this thesis, was given (Hiller et al 2009; Tugarinov et al 2005a). It is sufficient here to say that 4D NOESY sequences utilize only large one-bond scalar couplings ($^{1}J_{\mathrm{CH}}$= 120-150 Hz, $^{1}J_{\mathrm{HN}}$= 90-95 Hz) for coherence transfers. In effect, its performance does not degrade dramatically with increased molecular size. This suggests a perfect *niche* for SSA, which natively generates high quality 4D spectra, and avoids a somewhat troublesome 3D processing of 4D NUS data, necessary for acceptable convergence of MDD model (Tugarinov et al 2005a).

4.2.4. CHD-isotope filtered 4D NOESY spectra

Stereospecific assignments of prochiral protons in methylene (CH_2) groups and identification of pro-S/pro-R methyl groups (of Leu and Val) are essential for determination of high-precision NMR structures of proteins (Driscoll et al 1989; Güntert et al 1989; Kainosho et al 2006). In their absence, stereospecific assignments are deduced during iterative structure calculations by swapping methylene protons (Folmer et al 1997), or can be corrected manually.

Alternatively, pseudoatoms can be introduced, which leads to uncertainties in distance restraints and consequently to less precise structures (Güntert et al 1989). A severe problem with interpretation of NOESY spectra arises due to effective spin diffusion between methylene protons that can average observed cross-peaks so that remote peaks are indistinguishabe from direct ones. Effectively, little stereospecific information is available in NOESY spectra recorded using usual mixing times (100-150 ms). A remedy for this problem has been recently suggested by F. Ferrage and co-workers (Ecole Normale Supérieure, Paris). Their approach requires the preparation of partially randomly deuterated sample, which is far more accessible than the alternative stereo-array isotope labelling (SAIL) approach (Kainosho et al 2006). Two of four isotopomers provide NOE unaffected by spin diffusion as explained in Fig. 4.20. To filter out desired isotopomers, a CHD multiplicity filter was incorporated to *constant-time* 3D NOESY-^{13}C-HSQC experiment. The use of H(C)CO sequence (Yamazaki et al 1993) instead of *ct*-HSQC was suggested to restrict the information to sidechain protons of Asp, Asn, Glu and Gly residues, and also significantly improve resolution (see Fig.4.21 for magnetization transfer scheme). These aminoacids are prevalent in active sites of enzymes since their functional groups are frequently involved in specific hydrogen bonding and metal-ion-ligand interactions. The author was invited to join the project and contribute with four-dimensional extensions of CHD-edited NOESY experiments. This was accomplished by

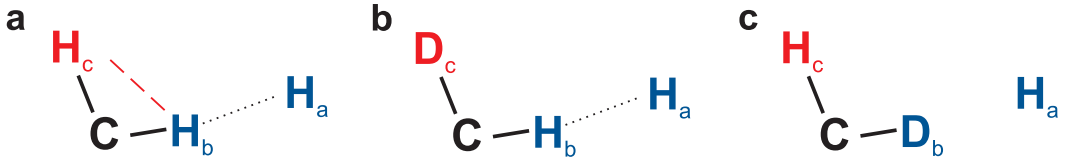
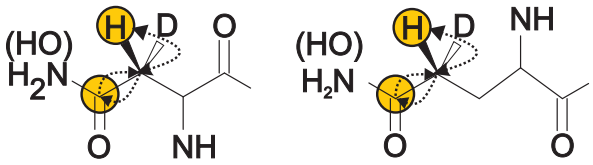


Fig. 4.20 Deleterious effects of spin diffusion (---) in CH_2 moieties which can lead to undesired remote NOE contacts between distant H_a and H_c spins (*a*). CHD isotopomers in fractionally ^{2}H-labelled proteins are unaffected by spin diffusion and exhibit direct correlation of peak volumes with interproton distance (*b*). However, the signal loss is incurred as H_a–H_b contact is missing for the second isotopomer (*c*).

Fig. 4.21 Magnetization transfers (*dotted arrows*) in sidechains of Asn (Asp) and Gln (Glu) during 2D H(C)CO sequence which is incorporated to 4D NOESY experiment after NOE mixing time (Fig. 4.22). Evolved spins are highlighted in *yellow*.



concatenation of optimized HMQC block, NOE mixing period and HSQC or H(C)CO for 4D ^{13}C-HMQC-NOESY-^{13}HSQC and 4D ^{13}C-HMQC-NOESY-H(C)CO, respectively. The pulse sequences are shown in Fig. 4.22. Noteworthy, to avoid *constant-time* ^{13}C evolution period, which could be detrimental for sensitivity for larger proteins, CHD-filter was incorporated into initial HMQC block by insertion of extra delays of total duration of only J_{CH}^{-1}, during which $^1J_{CH}$ is free to evolve. On the contrary, no additional sequence elements are necessary for H(C)CO-type NOESY since CH_2 isotopomers are suppressed by ^{1}H decoupling. Therefore, both 4D experiments can be recorded using minimal 4-step phase cycling which results in $\sqrt{2}$ gain in S/A ratio, compared to originally proposed 3D CHD-filtered NOESY sequences.

The proposed experiments were applied to a Ca^{2+}-loaded sample of the P43G mutant of calbindin D_{9k} (76 a.a., 8.8 kDa) with uniform ^{15}N,^{13}C- and random 50% ^{2}H-labelling. In the case of 4D C,C-edited NOESY the evolution of ^{13}C-^{13}C couplings limited ^{13}C evolution times to 6 (t_2) and 10 ms (t_3) during intital CHD-filtered HMQC and final HSQC parts. Despite significant experimental time of 149 h, only a fraction of 2% of points on a 84 × 66 × 110 Cartesian grid were sampled, which shows the resolution enhancement compared to conventional acquisition. For demonstrative purposes, β-methylene peaks of Leu25 and Leu28 from the characteristic and reasonably resolved fragment of ^{13}C-HSQC were selected (Fig. 4.23). Since CHD-filter is incorporated in the HMQC part which labels *NOE source* protons, the correlation planes of *destination* protons, i. e. ω_3(^{13}C)–ω_4(^{1}H, ditectly detected), are compared in Fig. 4.24. Besides excellent resolution that greatly simplifies the assignment of cross-peaks, there are significant differences in observed correlations for Hβ_2 and Hβ_3 *source* protons. The differences fully reflect the distinct cross-relaxation pathways of these protons in CHD (or CDH) isotopomers, as verified by comparison with crystal structure of calbindin D_{9k} (panels *c, d*). E.g., the cross-peak to Val61 Hβ is present only in Leu23 Hβ_3 plane, whereas intraresidue peak to Hα could be observed only for Leu23 Hβ_2. This information can be utilized in both ways: to assign Leu H$\beta_{2,3}$ protons stereospecifically basing on intraresidue contacts, or to introduce stereospecific long-range distance restraints for structure calculation.

The apparent S/N of 4D NUS C,C-edited NOESY was greatly enhanced by SSA. The average effective noise reduction, measured at the ^{1}H (ω_4) positions 363 diagonal peaks was 85.8%, which corresponds to reduction of 7.34 to 1.05 w. r. t. baseline (thermal) noise.

Even better spectral resolution was obtained with H(C)CO-based 4D sequence. The basis

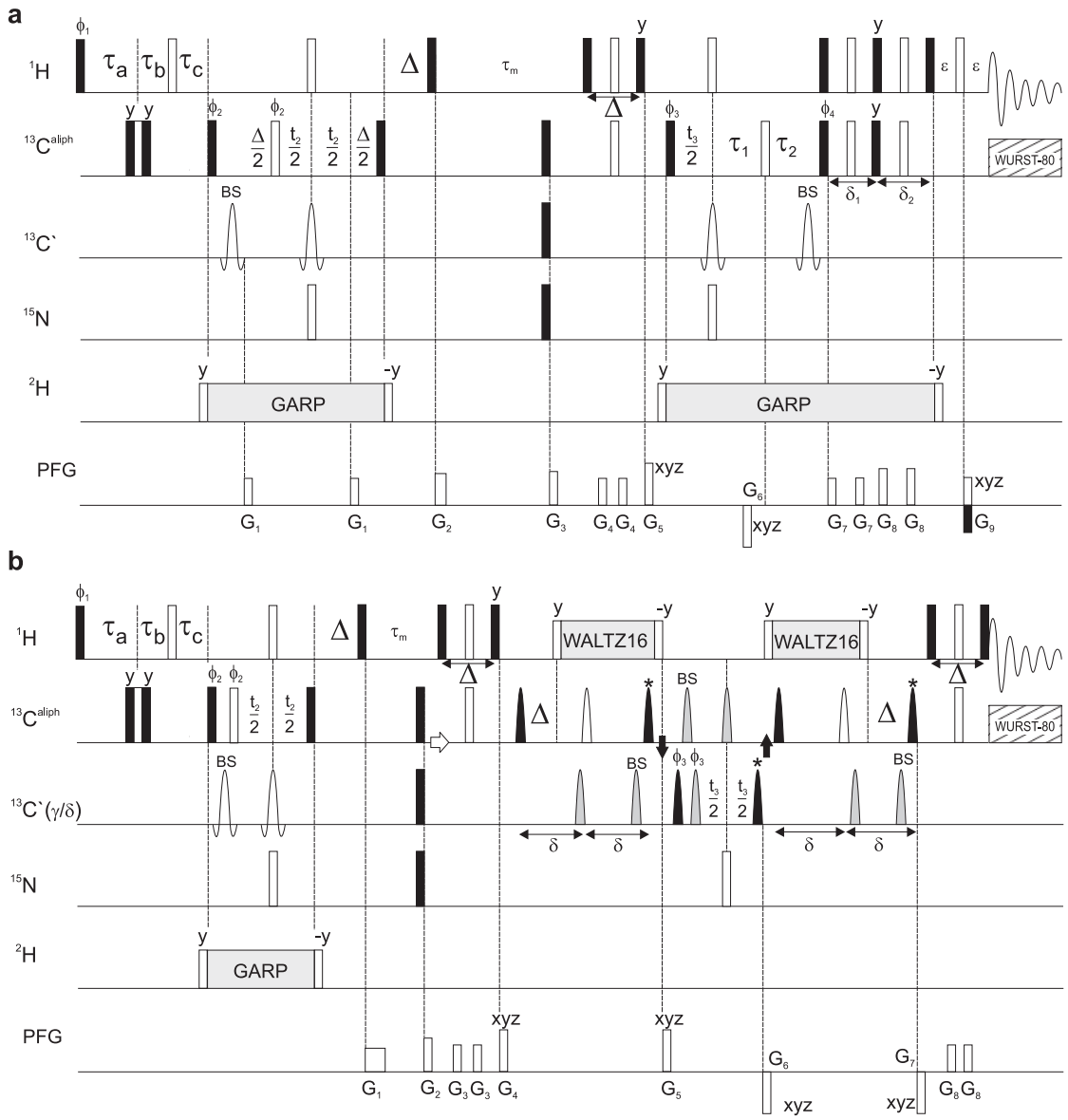


Fig. 4.22 Pulse sequence schemes for CHD-filtered (*a*) 4D ^{13}C-HMQC-NOESY-^{13}C-SE-HSQC and (*b*) 4D ^{13}C-HMQC-NOESY-H(C)CO. In both sequences ^{2}H-decoupling is employed whenever transverse aliphatic ^{13}C magnetization is present. CHD filter is implemented in seqence *a* by insertion of a pair of $\frac{1}{2}\Delta$ delays, $\Delta = (J_{\mathrm{CH}})^{-1}$, during ^{13}C evolution (t_2) in the initial HMQC. In the H(C)CO part of sequence *b* ^{1}H-decoupling supresses magnetization originating from CH_2 groups, which are in antiphase to ^{1}H after a delay of Δ after excitation of β/γ ^{13}C magnetization. Band selective Q3 and Q5 pulses are employed to select only sidechain resonances and suppress backbone HαCαC' groups. For both sequences NOE mixing time of τ_m=200 ms was used which was possible due to inhibited spin diffusion.

spectrum is reasonably immune to overcrowding due to selection of only Hβ and Hγ protons of Asp, Asn, Glu, Gln and utilization of well dispersed carbonyl (Cγ or Cδ) chem. shifts (Fig. 4.25). It is also noteworthy that only 3 out of 19 Q,E,N,D residues in calbindin D_{9k} have degenerate H$\beta_{2,3}$ (or H$\gamma_{2,3}$). C' (t_3) was evolved to a moderate value of $t_{3,max}$=32 ms, which suffices to resolve remaining 32 peaks in 2D H(C)CO spectrum. Max. evol. times in ω_1 (^{1}H) and ω_2 (^{13}C) were set to 20 and 12 ms, respectively, yielding relative sampling density

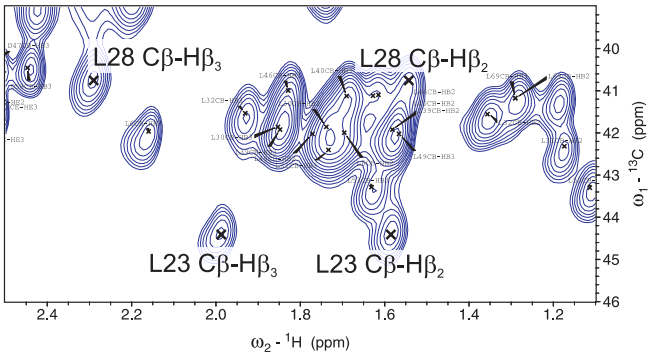


Fig. 4.23 Characteristic Leu Hβ region of ^{13}C-HSQC of calbindin D_{9k}. Diastereotopic resonances of Leu23 and Leu28, referred to in Fig. 4.24, are labelled. Noteworthy is the significant spectral crowding despite the relatively small size of the protein (76 a.a.).

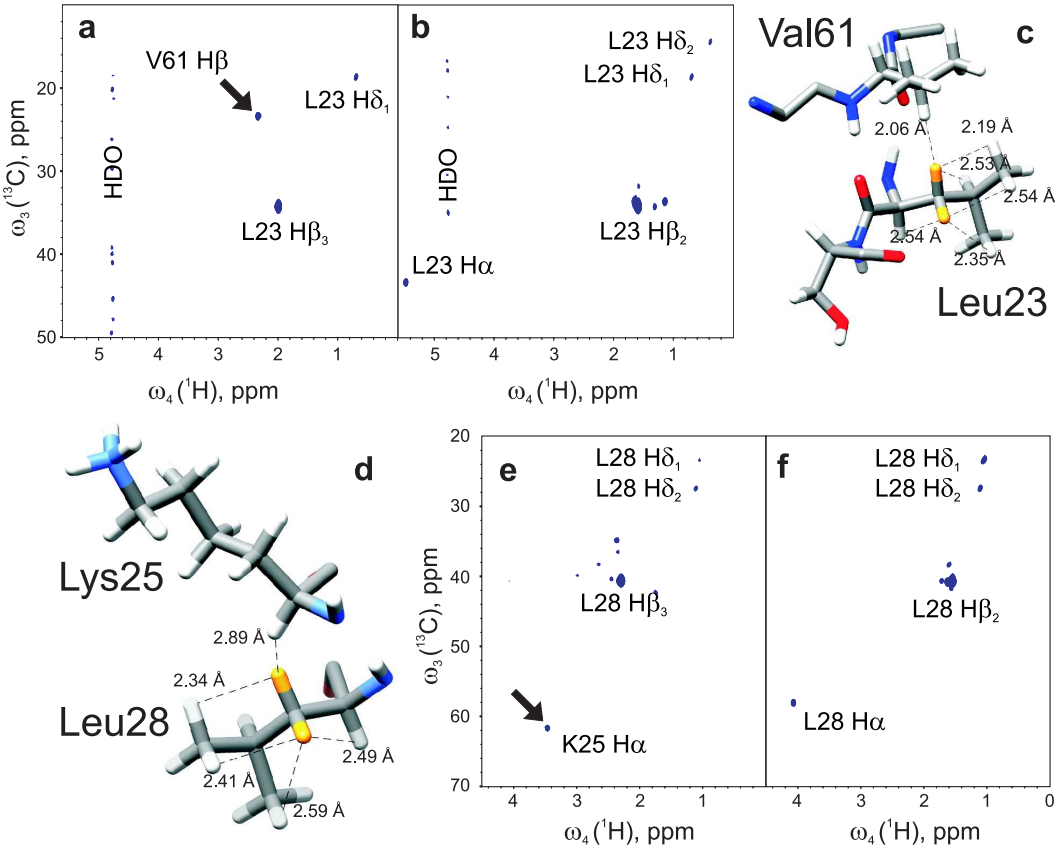


Fig. 4.24 2D ω_3 (^{13}C) / ω_4 (^{1}H) cross-sections from 4D C,C-edited HMQC-CHD-NOESY-HSQC spectrum for β-methylene groups of Leu23 (*a, b*) and Leu28 (*e, f*). The corresponding fragments of X-ray derived structure of calbindin D_{9k} (PDB code 4icb) are shown in *c* and *d*. The difference in intensity of observed intra- and inter-residual (marked with arrows) cross-peaks fully reflects distinct local environments of diastereotopic β protons.

of only 0.3%. The ω_1–ω_2 cross-sections, which illustrate correlations of Asp54 Hβ_2 and Hβ_3 to *NOE source* protons, are shown in Fig. 4.26. Again, the efficiency and relevance of CHD-filter is confirmed by substantial differences in the observed cross-relaxation networks for these diastereotopic β protons. The application of SSA resulted in average noise reduction of 37% (from 1.6 to 1.0 w.r.t. thermal noise) as measured at the ^{1}H (ω_4) positions of 39 autocorrelation peaks. The moderate gains in sensitivity reflect the smaller contribution of NUS artefacts than in the case of more sensitive C,C-edited NOESY experiment.

The measurements were carried out in the framework of Access to Infrastructure and Joint

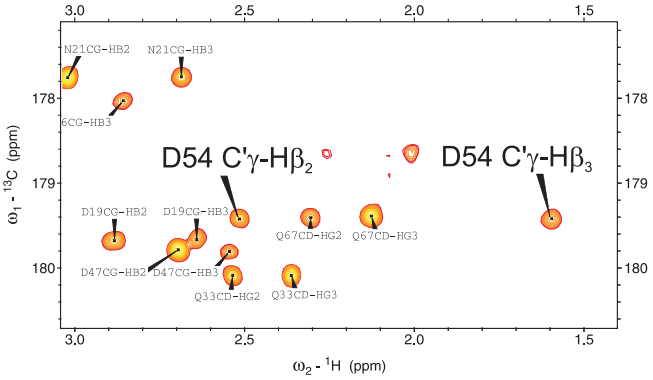


Fig. 4.25 Excerpt of the selective 2D H(C)CO spectrum of calbindin D_{9k} which shows sidechain carboxyl/carbonyl groups of Asp, Asn, Glu and Gln. Noteworthy is the excellent resolution in comparison to ^{13}C-HSQC (Fig. 4.23) due to relatively low number of resonances, good dispersion in sidechain carbonyl $^{13}C^{\gamma/\delta}$ dimension (ω_1) and large differences of diastereotopic Hβ or Hγ chem. shifts. Peaks of β protons of Asp54, referred to in Fig. 4.26, are highlighted.

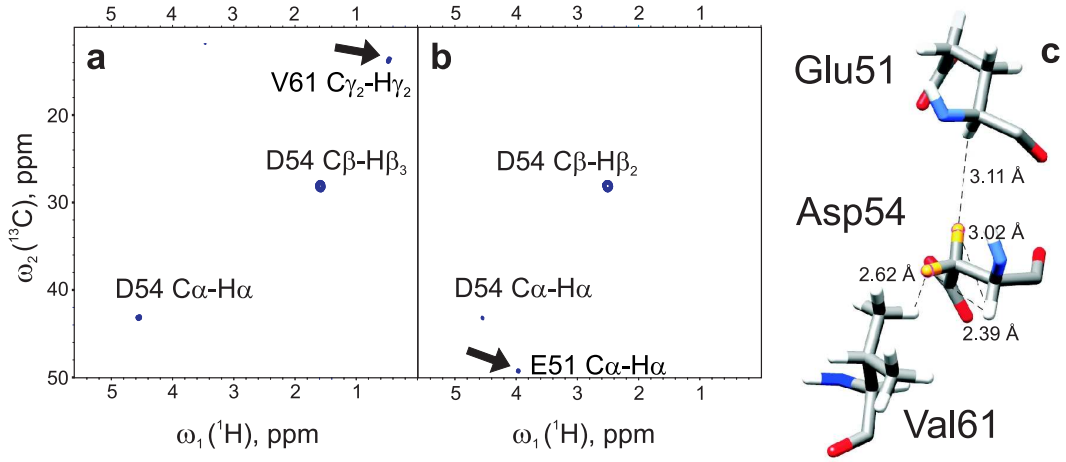


Fig. 4.26 ^{1}H (ω_1)–^{13}C (ω_2) cross-sections from 4D C-HMQC-NOESY-CHD-H(C)CO spectrum for β-methylene protons of Asp54 (*a, b*).The most relevant interresidual cross-peaks are marked with arrows. The observed ^{1}H-^{1}H NOE distance constraints are superimposed to X-ray derived structure (PDB code 4icb) using dashed lines and distance estimations (*c*).

Research Activity in the EAST-NMR project. The results will be published in the forthcoming months (*Nuclear Overhauser Spectroscopy of Chiral Methylene Groups*, R. Augustyniak, J. Stanek, H. Colaux, G. Bodenhausen, W. Koźmiński, T. Herrmann, F. Ferrage).

It should be emphasised that CHD-filtered 4D NOESY experiments have potential applications to partially deuterated large proteins due to attenuated transverse relaxation of involved ^{1}H and ^{13}C nuclei, and superior spectral resolution. Additionally, longer NOE mixing times are possible due to minimized spin diffusion effects which can lead to detection of long-distance restraints. Selection of CHD isotopomers not only enhances resolution and sensitivity (Lemaster 1990; Gardner and Kay 1998) but also enables more quantitative interpretation of NOESY cross-peak volumes (Lemaster 1987). The latter is also possible for mixtures of isotopomers (Nietlispach et al 1996), but requires a particularly tedious calibration which is considered a major bottleneck in practical use of NOESY in partially deuterated proteins (Cavanagh et al 2007). However, for deuteration level d the dilution of ^{1}H spins and selection of isotopomers result in signal of $1-d$, $d(1-d)$, and $d^2(1-d)$ of initial intensity for CH, CHD, CHD_2 groups, respectively. Attenuated transverse relaxation can compensate for this signal loss up to deuteration levels of 50-70% (Gardner and Kay 1998).

Table 4.3 Double-edited diagonal-free NOESY experiments.

Destination magnatization	*Source magnetization*		
	H^N	H^{ali}	H^{aro}
H^N	Diercks et al (2010) Werner-Allen et al (2010) Wu et al (2004)	Kay et al (1990a)	(Kay et al 1990a)[a]
H^{ali}	Hiller et al (2009)	Wu et al (2004) Wen et al (2012)	Stanek et al (2013a)
H^{aro}	–[b]	Stanek et al (2013a)	–[b]

[a]Requires broadband ^{13}C-HMQC
[b]Usually omitted or information available in 3D NOESY experiments

4.2.5. Diagonal-free symmetric and asymmetric NOESY spectra

Four-dimensional NOESY sequences incorporate two HMQC/HSQC blocks for heteronuclear-editing. It is then possible to selectively manipulate magnetization to distinguish between different kinds of possible NOE contacts. They are summarized in Table 4.3. Notably, it is possible using various methods to suppress auto-correlation peaks. Originally, it was aimed to eliminate truncation artefacts that may obscure cross-peaks in the proximity of diagonal peaks, and to remove t_1-noise (Wu et al 2004; Diercks et al 2010). However, in the case of NUS 3D or 4D spectra it also removes the main source of reconstruction artefacts (Stanek et al 2012; Wen et al 2012).

4D ^{13}C(ali),^{13}C(ali)-HMQC-NOESY-HSQC(HMQC) As suggested by Wu et al (2004), we have adapted C,C-edited NOESY experiment to suppress diagonal peaks by subtraction of signals from two interleaved measurements (see sequence in Fig. 4.27): in the standard one the magnetization flow is as follows

$$\underbrace{^1\mathrm{H}\,(\mathrm{SCT}\ t_1) \xrightarrow{J_{CH}} {}^{13}\mathrm{C}\ (t_2) \xrightarrow{J_{CH}} {}^1\mathrm{H}_z}_{\text{HMQC}} \xrightarrow{\tau_{NOE}} \underbrace{{}^1\mathrm{H}_z \xrightarrow{J_{CH}} \mathrm{H_zC_z}\ (t_3) \xrightarrow{J_{CH}} {}^1\mathrm{H}(t_4)}_{\text{HSQC}}$$

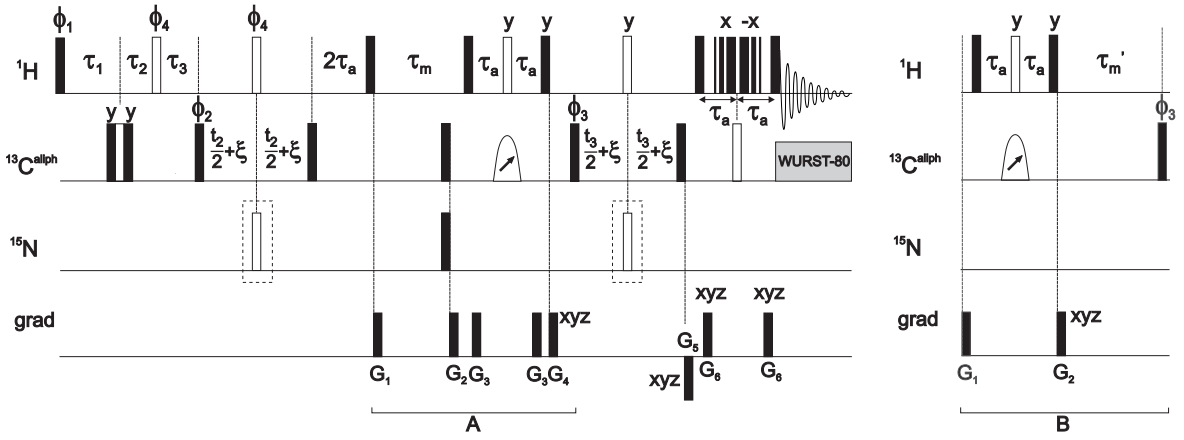


Fig. 4.27 Pulse scheme of the 4D ^{13}C-HMQC-NOESY-^{13}C-HSQC. Rough diagonal suppression is obtained by interleaved acquisition using the conventional (*left panel*) and modified sequence, in which the part (A) is replaced by scheme (B, *right panel*) to leave only auto-correlated signals. Subsequently, two FIDs have to be subtracted.

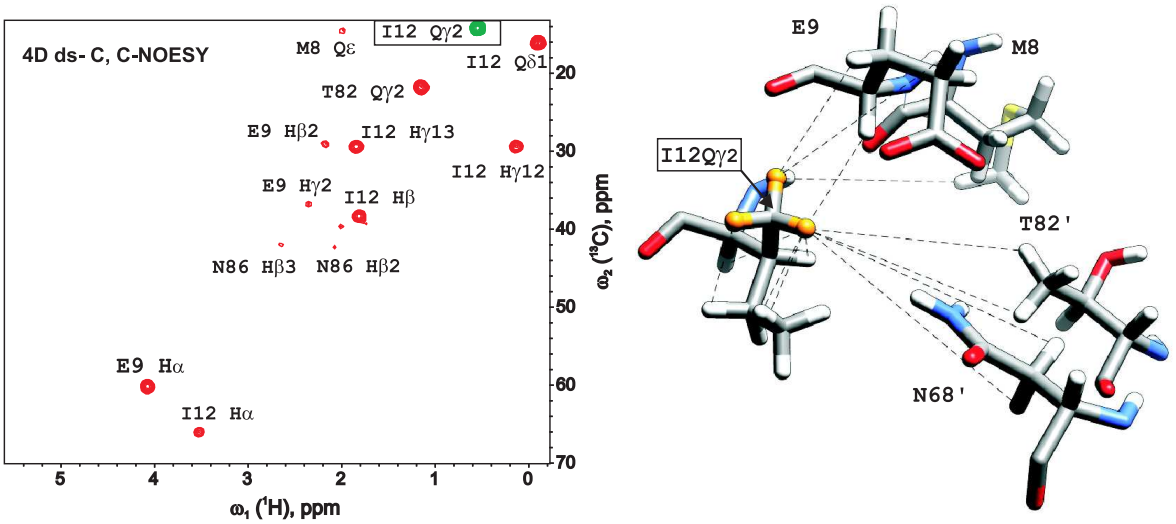


Fig. 4.28 A representative plane from diagonal-suppressed 4D C,C-edited NOESY of S100A1 (E32Q) protein. A slightly negative peak of Ile12 Qγ_2 is plotted in *green* and enclosed in a frame. All observed NOE cross-peaks are labelled, verified against conventional 3D ^{13}C-edited NOESY (*not shown*) and structural model (*right panel*), obtained previously from conventional NMR experiments (PDB code *2LHL*).

while in the modified version, the NOE mixing period is shifted after INEPT in the final HSQC. As a result, only auto-correlation peaks are refocused for acquisition:

$$\underbrace{{}^{1}\mathrm{H}\,(\mathrm{SCT}\ t_1) \xrightarrow{J_{CH}} {}^{13}\mathrm{C}\ (t_2) \xrightarrow{J_{CH}} {}^{1}\mathrm{H}_z}_{HMQC} \xrightarrow{J_{CH}} \mathrm{H_zC_z} \xrightarrow{\tau_{NOE}} \begin{matrix}\mathrm{H_zC_z}\ (t_3) \\ \mathrm{H'_zC_z}\ (t_3)\end{matrix} \begin{matrix}\xrightarrow{J_{CH}} \\ \xrightarrow[J_{CH'}=0]{}\end{matrix} \begin{matrix}{}^{1}\mathrm{H}(t_4) \\ \times\end{matrix}$$

The experiment was applied to an aforementioned S100A1 (E32Q) protein sample. A representative plane, shown in Fig. 4.28, exhibits a slightly negative weak peak at the diagonal position, however, this can be regulated by adjustment of signal multiplier upon subtraction of interleaved FIDs. To ensure equal experimental time the interleaved experiment was performed with 50% of points (3200) used for the above described usual C,C-NOESY. As demonstrated in Fig. 4.29*a* a tremendous reduction of effective noise in the z.a. FT spectrum is obtained which should be attributed to dramatically decreased dynamic range of signals. Nevertheless, the artefacts stemming from residual diagonal signals as well as the strongest cross-peaks are still above thermal noise (Fig. 4.29*b*). The noise in the usual and diagonal-suppressed (ds-) SSA-reconstructed spectra are compared in Fig. 4.29c. Although SSA artefacts are removed completely in ds- spectrum, this comes at the price of 50% of signal loss. This is because thermal noise is accumulated during entire experiment while useful cross-correlated signal only for odd transients. The decrease of S/N ratio is then larger than for uniform sampling, where it is only $\sqrt{2}$ for reasonably off-diagonal peaks (Wu et al 2004). In the case of NUS, the artefacts originating from diagonal peaks are *non-local*, and one-to-one subtraction is necessary for entire spectrum, i.e. also off-diagonal regions.

Although diagonal-suppression seems useful (see demanding application in Wen et al 2012), for typical applications the performance of SSA is sufficient to provide better S/N in the same experimental time: as shown in Table 4.2 the average S/N obtained for the conventional NUS C,C-NOESY is only 14% lower than, hypotetically, in absence of NUS artefacts. In other words, 50% signal loss in ds-spectra offsets most benefits from decreased artefact level. The

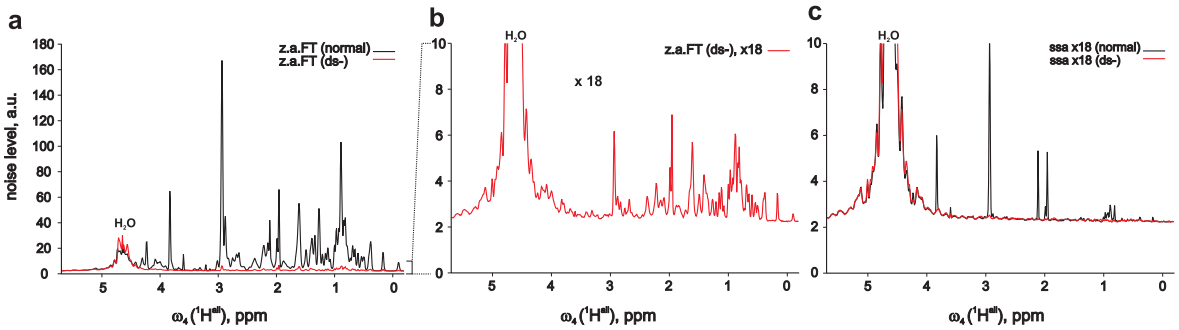


Fig. 4.29 Noise level in conventional (*black curves*) and diagonal-suppressed (*red curves*) 4D C,C-edited NOESY spectra recorded in the same time. (*a*) Noise in the z.a. FT spectra: conventional (*upper curve*) and diagonal suppressed (*bottom curve*), (*b*) 18-fold magnification of *a*, (*c*) Noise level in the SSA spectra, plotted in the same scale as *b*.

possibility to boost acquisition of ds-C,C-edited NOESY spectra by utilization of symmetry of purely auto-correlation signals is currently investigated[5].

4D ^{15}N-HSQC-NOESY-^{15}N-HSQC Diagonal suppression in N,N-edited spectra can be performed employing *orthogonal spin-state-selection* (oS^3) in TROSY-NOESY-TROSY experiment (Diercks et al 2010), which was recently exploited for the NUS acquisition of high-resolution 4D spectra (Werner-Allen et al 2010). However, 2-fold decrease in S/N of cross-peaks is incurred due to equal efficiency of ^{15}N spin state preservation and inversion during stochastic NOE transfer:

$$\mathrm{H_z\,N^{\alpha}} \quad \rightarrow \underbrace{\frac{1}{2}\mathrm{H'_z\,N^{\alpha}}}_{\text{discarded}} + \frac{1}{2}\mathrm{H'_z\,N^{\beta}}$$

Additionally, the pulse sequence is extended for necessary substitution of INEPT by a reversed ST2-PT block after NOE mixing period (Diercks et al 2010). For proteins with negligible TROSY effect the method of Wu et al (2004) can be applied, however, the potential benefits of diminished NUS artefacts are supposedly even smaller than for C,C-edited NOESY because of lower intrinsic sensitivity.

4D ^{13}C-HMQC-NOESY-^{15}N-HSQC In contrast to *symmetric* NOESY experiments, C,N-edited NOESY (Kay et al 1990a) inherently lacks auto-correlation signals and is therefore a perfect candidate for NUS (Stanek et al 2012). The application of SSA is usually necessary to retain optimal sensitivity, however, typical gains of S/N do not exceed 20-30% (see Fig. 4.30). Despite no signal is sacrificed to obtain diagonal-free spectrum, the increased dimensionality and sequence elongation considerably affect sensitivity. For the case of Lipase A, this has lead to ca. 48% fewer cross-peaks w.r.t. 3D N-edited NOESY recoded in the same time, and induced 54% loss of distance restraints. Hiller et al (2009) also noticed the larger fractions of missing peaks than for C,C-NOESY when compared to the number of distance restraints expected on the basis of crystal structure of 283-residue protein. This confirms that 4D C,N-edited experiment should be used only for very complex proteins.

4D ^{13}C(ali),^{13}C(aro)-HMQC-NOESY-HSQC Recently, the author proposed an asymmetric C(ali),C(aro)-edited NOESY experiment which in fact utilizes the same principle as

[5]Iuventus Plus, a grant No. IP2012 057872 (2013-2014) by Polish Ministry of Science and Education

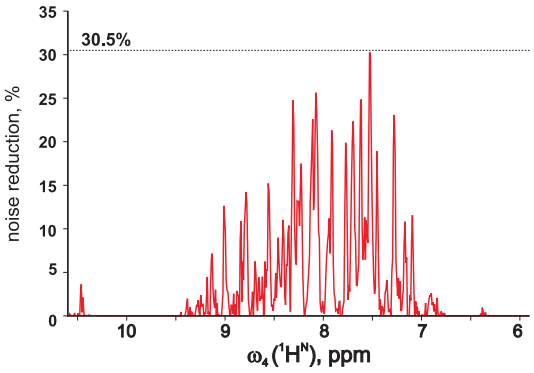


Fig. 4.30 Noise reduction in the 4D C,N-edited NOESY spectrum of S100A1 (E32Q). Sensitivity gains are moderate since only cross-peaks contribute to NUS artefacts. Thus, C,N-NOESY is in the intermediate regime between *sampling-* and *sensitivity-limited* experiments.

C,N-NOESY, namely the possibility of selective manipulation of alipahtic and aromatic ^{13}C spins. This is possible by use of only 2 or 3 shaped pulses as shown in Fig. 4.31.

Again, the experiment was performed for a sample of S100A1 (E32Q). It was demonstrated that the experiment provides a number of relevant long-range distance restraints since aromatic residues are usually hidden in the innards of proteins (Stanek et al 2013a). In this case, application of SSA turned out pointless as NUS artefacts did not apparently increase the noise level as measured in z.a. FT spectrum (Fig. 4.32). Noteworthy, the same principle of

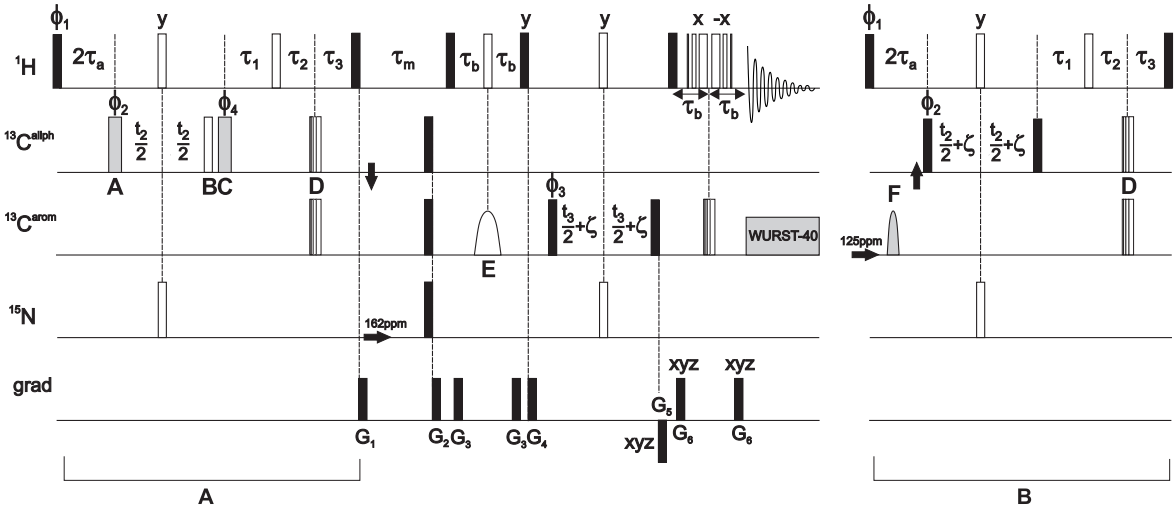


Fig. 4.31 Pulse scheme of 4D ^{13}C(ali)-HMQC-NOESY-^{13}C(aro)-HSQC. The selectivity of coherence transfer is obtained by phase-cycled selective ^{13}C (ali) 90° pulses (A, C) in HMQC, and shaped inversion pulse (E) with the following PFG G_4 in ^{13}C(aro)-HSQC. The alternative implementation of ^{13}C(ali)-HMQC, which employs a single ^{13}C(aro) inversion pulse (F), is also shown (*right panel*).

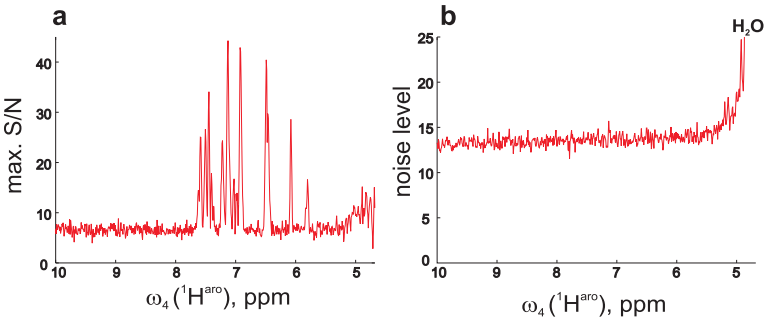


Fig. 4.32 Since 4D ^{13}C(ali),^{13}C(aro)-edited NOESY contains only cross-peaks of at most moderate S/N (panel *a*), the contribution of NUS artefacts to z.a. FT spectrum is negligible (panel *b*). The experiment is thus *sensitivity-limited* and no improvement can be obtained using SSA.

asymmetric C,C-NOESY was recently proposed for high-resolution ribose-aromatic NOESY which significantly enhances the sequential assignment of RNAs (Stanek et al 2013b).

In conclusion, at the identical sample conditions the NUS artefacts have the largest influence on ds-^{13}C,^{13}C-edited NOESY, and smaller or negligible on ^{13}C,^{15}N- and ^{13}C(ali),^{13}C(aro)-edited experiments. Nevertheless, the deleterious effects of NUS artefacts can be present in all these spectra, e.g. at higher B_0, and this has to be verified empirically for each case.

4.3. Cross-correlated relaxation measurements

For decades the primary source of structural constraints were interatomic distances derived from NOE (Overhauser 1953; Solomon 1955) and dihedral angles derived from three-bond scalar couplings (Karplus 1963). These approaches have two major disadvantages: NOE rapidly decreases with r^{-6} and multiparametric calibration of empirical Karplus curve have to be properly done for ^3J. In the seminal work Reif et al (1997) demonstrated that another quantity, namely cross-correlated relaxation rate, provides a direct correlation with protein backbone dihedral angles (Fig. 4.33). In the following years, a variety of experiments were

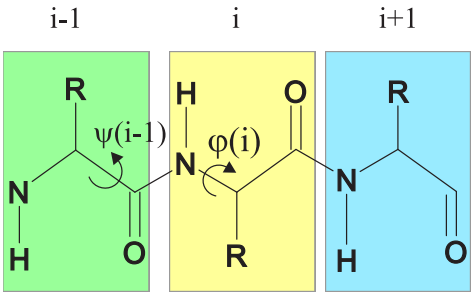


Fig. 4.33 Scheme of the peptide chain with the definition of dihedral angles φ and ψ.

proposed to measure both dipolar-dipolar (DD)[6] and dipolar-chemical shift anisotropy (DD-CSA) cross-correlation rates. In the slow tumbling regime, the expressions for a rigid and isotropically tumbling molecule are as follows (Goldman 1984):

$$\Gamma_{i,j} = \left(\frac{\hbar\mu_0}{4\pi}\right)^2 \frac{2}{5}\frac{\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}}{r^3_{\mathrm{HN}}}\frac{\gamma_{\mathrm{H}}\gamma_{\mathrm{C}}}{r^3_{\mathrm{HC}}} \times \frac{1}{2}(3\cos^2\theta - 1)S^2\tau_C \tag{4.3}$$

for dipoles i (say H–N) and j (e.g., C$^\alpha$–H$^\alpha$) that form a dihedral angle θ, and

$$\Gamma_{i,k} = \frac{\hbar\mu_0}{4\pi}\frac{2}{5}\frac{\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}}{r^3_{\mathrm{HN}}}\frac{2\gamma_k B_0}{3} \times f(\boldsymbol{\sigma_k})S^2\tau_C \tag{4.4}$$

for cross-correlation of dipole i (say H–N) and CSA tensor $f(\boldsymbol{\sigma_k})$. In the latter case $f(\boldsymbol{\sigma_k})$ denotes the function of projections of dipole i unit vector (terms $\frac{1}{2}(3\cos^2\theta_p - 1)$) onto principal axes of CSA tensor k:

$$f(\boldsymbol{\sigma_k}) = \frac{1}{2}\left(\sigma_x(3\cos^2\theta_x - 1) + \sigma_y(3\cos^2\theta_y - 1) + \sigma_z(3\cos^2\theta_z - 1)\right) \tag{4.5}$$

and thus, embeds structural information. In the above formulas S^2 is the generalized Lipari-Szabo order parameter (*local rigidity*) and τ_C stands for overall correlation time. Interestingly, DD CCR is field-independent, while DD-CSA scales linearly with static field B_0 (or ω_k). Usually, standard bond lenghts, the planarity of peptide bond and C' CSA tensor of

[6] Although the correct abbreviation for dipolar-dipolar interference is DD-DD, most commonly DD is used

$(\sigma_{xx}, \sigma_{yy}, \sigma_{zz})$=(244, 178, 90) ppm, with x an y axes located in the peptide plane and y axis tilted by 8° with respect to the carbonyl bond, are assumed. This allows to recast directional cosines onto backbone dihedral angles φ and ψ for available DD cross-correlations (Reif et al 1997; Pelupessy et al 1999; Chiarparin et al 2000):

$$\begin{array}{lcl} \cos\theta(\mathrm{C}^\alpha\mathrm{H}^\alpha\ (i-1),\ \mathrm{NH}^\mathrm{N}\ (i)) & \approx & 0.163\ +0.819\ \cos(\psi(i)-120^\circ) \\ \cos\theta(\mathrm{C}^\alpha\mathrm{H}^\alpha\ (i),\ \ \ \mathrm{NH}^\mathrm{N}\ (i)) & \approx & 0.163\ -0.819\ \cos(\varphi(i)-60^\circ) \\ \cos\theta(\mathrm{C}^\alpha\mathrm{H}^\alpha\ (i-1),\ \mathrm{C}^\alpha\mathrm{H}^\alpha\ (i)) & \approx & -0.106\ +0.894\ \cos(\psi(i-1)+\varphi(i)+180^\circ) \end{array} \tag{4.6}$$

and, similarly for inter- (Yang et al 1997, 1998) and intraresidual (Kloiber and Konrat 2000a) $\Gamma_{C',H\alpha C\alpha}$ DD-CSA correlations (for non-glycine residues):

$$\begin{array}{lcl} \cos\theta_x & = & \cos\theta\left(\sigma_x\ (i), \mathrm{C}^\alpha\mathrm{H}^\alpha\ \begin{smallmatrix}(i-1)\text{: }+\\ (i)\text{: }-\end{smallmatrix}\right) \approx -0.3095\ +0.3531\ \cos(\varphi(i)\pm 120^\circ) \\ \cos\theta_y & = & \cos\theta\left(\sigma_y\ (i), \mathrm{C}^\alpha\mathrm{H}^\alpha\ \begin{smallmatrix}(i-1)\text{: }+\\ (i)\text{: }-\end{smallmatrix}\right) \approx -0.1250\ -0.8740\ \cos(\varphi(i)\pm 120^\circ) \\ \cos\theta_z & = & \cos\theta\left(\sigma_z\ (i), \mathrm{C}^\alpha\mathrm{H}^\alpha\ \begin{smallmatrix}(i-1)\text{: }+\\ (i)\text{: }-\end{smallmatrix}\right) \approx -0.9426\ \sin(\varphi(i)\pm 120^\circ) \end{array} \tag{4.7}$$

The methodology has been well established and verified for well folded proteins. Additionally, to remove ambiguities arising from multiplicity of possible ψ, φ values matching Karplus-like curves, as well as to eliminate the role of internal dynamics, an automated protocol was proposed (Kloiber et al 2002).

CCR can be measured for multiple-quantum coherences of *arbitrary choice*, and the only practical limitation is the capability of coherence transfer between involved nuclei, for which usually scalar couplings are employed. This flexibility comes from the fact that CCR reflects the correlation of a relaxation process for independent (and, possibly, quite distant) nuclei creating MQ coherence. No strong physical interference of internal magnetic fields (such as in ^{15}N-TROSY) is thus required. Consequently, CCR provide valuable long-range (global) relative vector orientations, similarly to residual dipolar couplings. However, in contrast to RDCs no weak alignmnent medium is required, which is a significant advantage of CCR approach to structure determination. Instread, the internal reference frame is provided by a CSA tensor or another dipolar vector.

Nevertheless, nowadays CCR (similarly to ^3J) are not commonly measured in routine structure elucidation. This can be associated with increased crowding of the spectra (if multiplets are recorded), or necessity to collect a pair of specially edited spectra. Additionally, a relatively high S/N ratio must be obtained for a pair of peaks to provide meaningful CCR value.

Until recently, it was also presumed that measurement of CCR for IDPs is pointless, since it would provide statistically averaged values without proper interpretation. Moreover, the resolution of existing 3D experiments prevented these measurements due to extreme crowding in ^{15}N-HSQC spectra of IDPs. However, following the suggestion of Prof. R. Konrat, the original experiment for measurement of DD NH^N (i)–CSA C'(i) (Kloiber and Konrat 2000b) was redesigned to include additional C^α evolution period for necessary resolution enhancement (Fig. 4.34). The desired coherence pathways is as follows:

$$\begin{aligned} \mathrm{H}^\mathrm{N} \xrightarrow{J_\mathrm{NH}} {}^{15}\mathrm{N} \xrightarrow{{}^1J_{\mathrm{NC}\alpha}} {}^{13}\mathrm{C}^\alpha \xrightarrow{J_{\mathrm{C}'\mathrm{C}\alpha}} {}^{15}\mathrm{N}{}^{13}\mathrm{C}'\ (\mathrm{CT}\ t_1) \xrightarrow{J_{\mathrm{C}'\mathrm{C}\alpha}} \\ {}^{13}\mathrm{C}^\alpha\ (\mathrm{SCT}\ t_2) \xrightarrow{{}^1J_{\mathrm{NC}\alpha}} {}^{15}\mathrm{N}\ (\mathrm{SCT}\ t_3) \xrightarrow{J_\mathrm{NH}} \mathrm{H}^\mathrm{N}(t_4) \end{aligned}$$

The distinguishing feature of the experiment is the central MQ $^{15}\mathrm{N}^{13}\mathrm{C}'$ evolution period, which is designed to simultaneously evolve C' chemical shift and $^1J_\mathrm{NH}$ and refocus ^{15}N evolution as well as all undesired passive couplings (Fig. 4.35). Its crucial here that ZQ and

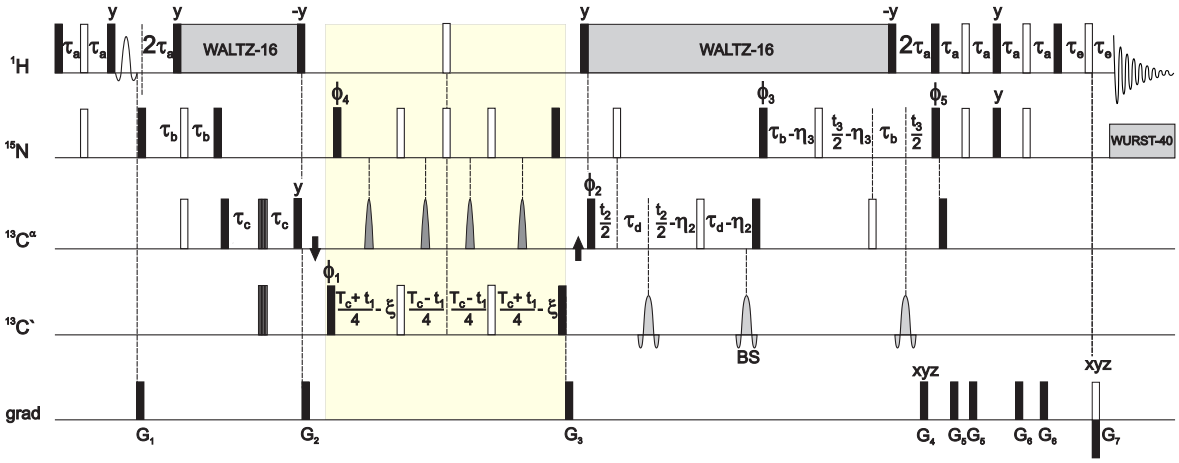
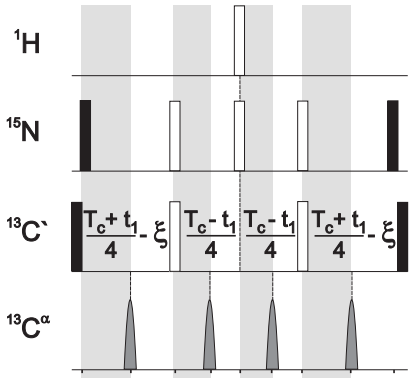


Fig. 4.34 Scheme of the 4D HNCACO-CCR(CSA C' – DD $H^N N$) experiment. The pulse sequence was adapted from 4D HNCACO (Zawadzka-Kazimierczuk et al 2010) by substitution of a C' evolution period with a CCR evolution block of length T_C (*highlighted*), which is dicussed in detail in Fig. 4.35. Fixed coherence transfer delays are: τ_a=2.69 ms $\approx (4J_{\mathrm{NH}})^{-1}$, τ_b=14 ms $\approx (8\,{}^1J_{\mathrm{NC}\alpha})^{-1}$, τ_c=3.15 ms $< (4J_{C'C\alpha})^{-1}$, τ_d = 4 ms $\approx (4J_{C'C\alpha})^{-1}$. C^α (t_2) and ^{15}N (t_3) evolutions are performed in the shared-time manner, with contractions $\eta_2 = \tau_d\, t_2/t_{2max}$ and $\eta_3 = \tau_b\, t_3/t_{3max}$, respectively. For other experimental details the Reader is referred to the supplement of the article (Stanek et al 2013c).

Fig. 4.35 Central part of the pulse sequence shown in 4.34, where DQ & ZQ $N_y C'_y$ coherence is stored for the constant period T_C. This (usually long) period is also used for C' frequency labelling (t_1) leading to superior resolution.
The analysis of relative signs and total effective evolution time of relevant chemical shifts, scalar coupling and cross-correlation rates is given below the pulse scheme.



Evolution of:									Total evolution time
$\Gamma_{NH,C'}$ (DD-CSA)	+	+	+	+	+	+	+	+	**T**
$\Gamma_{NH,N}$ (DD-CSA)	+	+	+	+	−	−	−	−	0
$\delta(^{13}C)$	+	+	−	−	−	−	+	+	$\mathbf{t_1}$
$\delta(^{15}N)$	+	+	−	−	+	+		−	0
$^1J_{NH}$	+	+	−	−	−	−	+	+	$\mathbf{t_1}$
$^1J_{NC'}$	not evolving (MQ)								0
$^1J_{NC\alpha}$, $^2J_{NC\alpha}$	+	−	+	−	+		+	−	0
$^1J_{C'C\alpha}$	+	−	+	−	−	+	+	−	0

DQ coherences are interconverted simultaneously (by π pulses for ^{15}N and C'), or along with a flip of H^N, which ensures the preservation of relative sign of $\Gamma_{C',HN}$ contribution to the relaxation rates. E.g., let us analyse the effect of ^{15}N and C' π pulses on ZQ,α and DQ,β

coherences:[7]

$$\begin{aligned}
\mathrm{DQ}_-,\alpha &\equiv (\alpha\beta\beta \leftrightarrow \alpha\alpha\alpha) \xrightarrow{\pi(^{15}\mathrm{N})} \mathrm{ZQ}_+,\alpha \quad (\alpha\alpha\beta \leftrightarrow \alpha\beta\alpha) \xrightarrow{\pi(^{13}\mathrm{C}')} (\alpha\alpha\alpha \leftrightarrow \alpha\beta\beta) \equiv \mathrm{DQ}_+,\alpha \\
\mathrm{ZQ}_+,\beta &\equiv (\beta\alpha\beta \leftrightarrow \beta\beta\alpha) \xrightarrow{\pi(^{15}\mathrm{N})} \mathrm{DQ}_-,\beta \quad (\beta\beta\beta \leftrightarrow \beta\alpha\alpha) \xrightarrow{\pi(^{13}\mathrm{C}')} (\beta\beta\alpha \leftrightarrow \beta\alpha\beta) \equiv \mathrm{ZQ}_-,\beta
\end{aligned}$$

To summarise, two pulses do not change the relevant coherence, except for the sign of chemical shift evolution (analogous conclusions can be drawn for ZQ,β and DQ,α). Next, the effect of simultaneous $\mathrm{H^N}$ and ^{15}N π pulses is analysed:

$$\begin{aligned}
\mathrm{DQ}_-,\alpha &\equiv (\alpha\beta\beta \leftrightarrow \alpha\alpha\alpha) \xrightarrow{\pi(^{15}\mathrm{N})} \mathrm{ZQ}_+,\alpha \quad (\alpha\alpha\beta \leftrightarrow \alpha\beta\alpha) \xrightarrow{\pi(\mathrm{H^N})} (\beta\alpha\beta \leftrightarrow \beta\beta\alpha) \equiv \mathrm{ZQ}_+,\beta \\
\mathrm{ZQ}_+,\beta &\equiv (\beta\alpha\beta \leftrightarrow \beta\beta\alpha) \xrightarrow{\pi(^{15}\mathrm{N})} \mathrm{DQ}_-,\beta \quad (\beta\beta\beta \leftrightarrow \beta\alpha\alpha) \xrightarrow{\pi(\mathrm{H^N})} (\alpha\beta\beta \leftrightarrow \alpha\alpha\alpha) \equiv \mathrm{DQ}_-,\alpha
\end{aligned}$$

In conclusion, these pulses interconvert DQ,α and ZQ,β (or, similarly, DQ,β and ZQ,α). Since the interconversion occurs exactly in the middle of T_C the effective relaxation rate is a an average for DQ,α and ZQ,β:

$$\exp\big(-\frac{1}{2}T_C\Gamma_{DQ,\alpha} - \frac{1}{2}T_C\Gamma_{ZQ,\beta}\big) = \exp\big(-T_C\underbrace{\frac{1}{2}(\Gamma_{DQ,\alpha}+\Gamma_{ZQ,\beta})}_{\Gamma_\alpha}\big) \tag{4.8}$$

In turn the relaxation rates for particular coherences are given by:[8]

$$\Gamma_{DQ,\alpha} = \Gamma_\mathrm{a} + \Gamma_{\mathrm{HN,N}} + \Gamma_{\mathrm{HN,C'}} + \Gamma_{\mathrm{N,C'}} \tag{4.9}$$

$$\Gamma_{ZQ,\beta} = \Gamma_\mathrm{a} - \Gamma_{\mathrm{HN,N}} + \Gamma_{\mathrm{HN,C'}} - \Gamma_{\mathrm{N,C'}} \tag{4.10}$$

$$\Gamma_{ZQ,\alpha} = \Gamma_\mathrm{a} + \Gamma_{\mathrm{HN,N}} - \Gamma_{\mathrm{HN,C'}} - \Gamma_{\mathrm{N,C'}} \tag{4.11}$$

$$\Gamma_{DQ,\beta} = \Gamma_\mathrm{a} - \Gamma_{\mathrm{HN,N}} - \Gamma_{\mathrm{HN,C'}} + \Gamma_{\mathrm{N,C'}} \tag{4.12}$$

Effectively, the coherences „$\mathrm{H^N}(\alpha)$" and „$\mathrm{H^N}(\beta)$" evolving as $\Omega_{C'} \pm \pi J_{\mathrm{NH}}$ undergo relaxation with averaged rates $\Gamma_{\alpha/\beta} = \Gamma_\mathrm{a} \pm \Gamma_{C',HN}$.

In the original paper Kloiber and Konrat (2000b) proved that $\Gamma_{C',NH}$, which is a function of both dihedral angles φ and ψ, is very sensitive to local conformation of the peptide chain (Fig. 4.36). In particular, it enables one to distinguish β-turns of type-I from type-II (Fig. 4.37).

It is known that IDPs show preference for particular transient secondary structure elements, which exhibits in deviations of chemical shifts from random coil averaged values. Thus, in the collaboration with the group of Prof. R. Konrat we decided to investigate whether a similar correlation could be implied from intra-residual $\Gamma_{C',NH}$. For this purpose, 4D HNCACO-CCR was applied to Chicken Brain Acid Soluble protein (cBASP1), an intrinsically disordered protein comprising 270 a. a. (Geist et al 2013b). The experiment was performed for two samples of cBASP1: at pH=2 and 6, for which the impact of pH on biological function and some standard NMR parameters was investigated elsewhere (Geist et al 2013a).

NUS with extraordinarily long evolution times was employed[9] to ensure high spectral resolution, and to fully utilize unusually long constant-time period used for evolution of CCR

[7] The notation xyz refers to a spin state of $\mathrm{H^N}$, ^{15}N and C' nuclei, respectively. ZQ,α refers to zero-quantum coherence with $\mathrm{H^N}(\alpha)$.

[8] Γ_a stands for auto-correlated relax. rate

[9] $t_{1max}(C') = 90\ ms$, $t_{2max}(C^\alpha) = 10\ ms$, $t_{3max}(N) = 50\ ms$

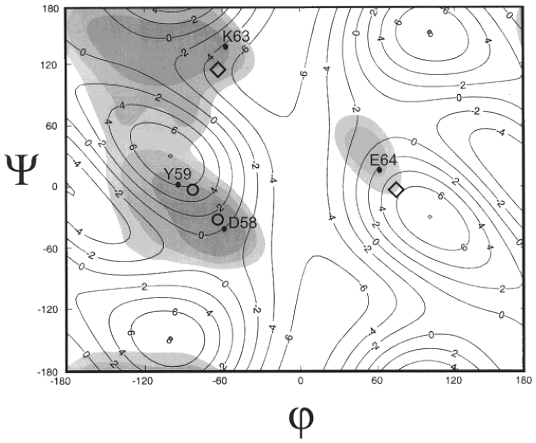


Fig. 4.36 The dependence of $\Gamma_{C',NH}$ on the dihedral angles φ and ψ in peptide backbone. Shaded regions correspond to allowed combinations of dihedral angles (Ramachandran plot). Indicated are the idealized angles (φ, ψ) for the central residues, $i+1$ and $i+2$, in type-I ($\circ$) and type-II ($\diamond$) β-turns. Additionally, experimentally determined geometries for Asp58, Tyr59, and Lys63 and Glu64, which form aforementioned turns in human ubiquitin, are labelled. Clearly, type-II β-turns can exhibit both positive and negative CCR values, while for type-I β-turns only positive values are allowed. Reprinted from Ref. (Kloiber and Konrat 2000b).

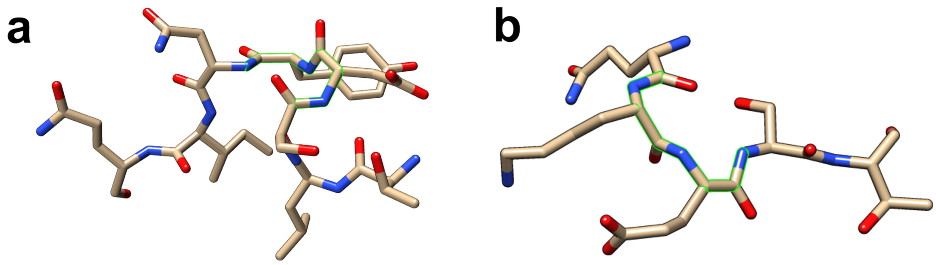


Fig. 4.37 Type I (a) and II (b) β-turns. Idealized dihedral angles for the central residues ($i+1$ and $i+2$) are as follows: $\varphi_{i+1} = -60°$, $\psi_{i+1} = -30°$, $\varphi_{i+2} = -90°$, $\psi_{i+2} = 0°$ (type I), and $\varphi_{i+1} = -60°$, $\psi_{i+1} = 120°$, $\varphi_{i+2} = 90°$, $\psi_{i+2} = 0°$ (type II). The structures correspond to β-turns formed by Ser57-Asp58-Tyr59-Asn60 (a) and Gln62-Lys63-Glu64-Ser65 (b) in human ubiquitin (based on structure deposited in PDB with accession code 1UBQ).

($T_C = 90\ ms$). This was only feasible owing to favourable relaxation properties of cBASP1. The relative sampling density was as small as only 0.24% and 0.36% for cBASP1 at pH=2 and 6, respectively. The spectra were reconstructed with SSA for its capability of faithful reproduction of relative peak amplitudes, which is crucial for determination of CCR rates. Noteworthy is the size of resulting spectra, $\approx$ 45 GB (digital resolution of 720 $\times$ 256 $\times$ 512 $\times$ 128 was set in dimensions $\omega_1 \ldots \omega_4$), which posed high requirements on the performance of processing software.

The representative C'-C^α planes for Thr108 are shown in Fig. 4.38. It is clearly demonstrated that additional C^α evolution period greatly enhances the spectral resolution comparing to to the original 3D experiment. C^α chemical shifts in IDPs tend to cluster in the regions specific for a given amino acid type, however, they are barely correlated to C', ^{15}N and H^Nchemical shifts. Consequently, the C^α dimension resolves the resonances in a few spectral subregions, which can be associated with an a. a. type. The inter-residual correlations present in Fig. 4.38 increase the spectral crowding, which is 4-fold larger than in 4D HNCOCA. These cross-peaks arise due to two-way coherence transfer, with the unwanted pathway utilizing $^2J_{\mathrm{NC}\alpha} \approx 4-7\ Hz$, and are usually of lower intensity. They do not provide any valuable structural information since they depend on *fixed* peptide bond geometry.

Despite extensive spectral crowding in 2D ^{15}N-HSQC spectrum, presence of both intra- and interresidual peaks, and J_{NH}-spitting of each resonance, it was possible to determine CCR rate for approx. 82% of BASP1 residues with amide protons in slow exchange regime

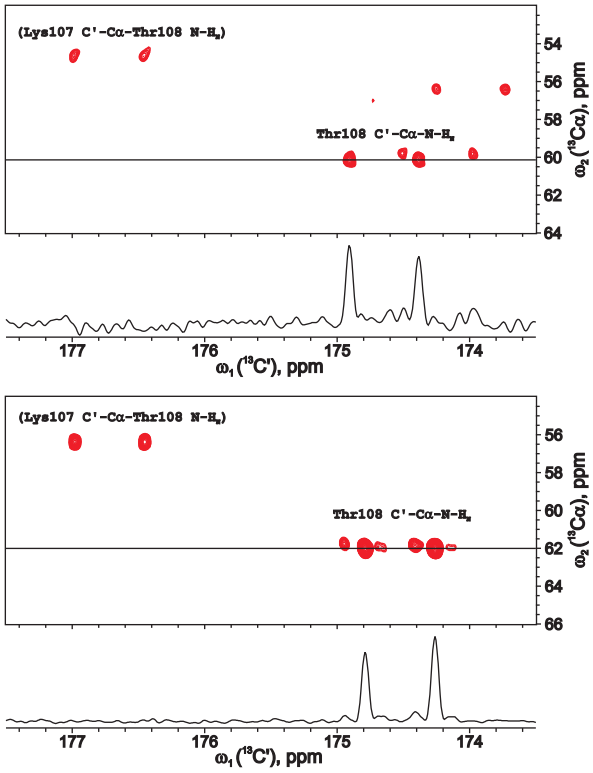


Fig. 4.38 C'(ω_1)–C$^\alpha$(ω_2) planes from 4D HNCACO-CCR(CSA C' – DD H^{N}N) spectrum of BASP1 at pH=2 (*top*) and pH=6 (*bottom panel*). The cross-sections contain intra-residual resonances of Thr108, and irrelevant inter-residual Lys107-Thr108 peaks (*labels in parentheses*). The provided 1D cross-sections along ω_1 (C') dimensions show the substantially distinct intensity of up- and downfield doublet lines. The relative intensity changes with pH dramatically, which corresponds to shift of $\Gamma_{C',HN}$ from negative to positive value.

at both pH=2 and 6. The CCR rates were calculated from the equation:

$$\Gamma_{C',HN} = \frac{1}{2T_C} \ln \frac{I_{\text{up}}}{I_{\text{down}}} \tag{4.13}$$

where I_{up} and I_{down} denote the amplitude of an up- and downfield component of J_{NH}-multiplet in ω_1, respectively. It is of particular importance here that the relative intensity of peaks is not disturbed by the application of decaying sampling density (see 2.4.2). The eq. 4.13 can be easily derived by noting that during the constant time period of length T_C the relevant multiplet components relax at different rates of $\Gamma_{\text{up/down}} = \Gamma_{\text{a}} \mp \Gamma_{C',HN}$. This results in attenuation of factor

$$I_{\text{up/down}}/I(t_0) = \exp(-T_C(\Gamma_a \mp \Gamma_{C',HN})) \tag{4.14}$$

which, by taking the side-by-side logarithm and subtraction of expressions for I_{up} and I_{down}, leads to the eq. 4.13.

It is quite interesting to perform error analysis for the expression from eq. 4.13. In the small-error approximation it yields

$$|\Delta\Gamma_{C',HN}| \approx \frac{1}{2T_C}\sigma_{\text{eff}}\left(\frac{1}{I_{\text{up}}} + \frac{1}{I_{\text{up}}}\right), \tag{4.15}$$

where σ_{eff} denotes the effective noise level which affects equally both lines of a doublet. After recalling eq. 4.14 one derives:

$$|\Delta\Gamma_{C',HN}| \approx \frac{1}{2T_C}\frac{\sigma_{\text{eff}}}{I_0}\exp(\Gamma_a T_C)(\exp(\Gamma_{C',HN}T_C) + \exp(-\Gamma_{C',HN}T_C)) \tag{4.16}$$

The analytical derivative of Eq. 4.16 is quite cumbersome, however, one can conclude that it reaches zero for T_C between $(\Gamma_a + \Gamma_{C',HN})^{-1}$ and $(\Gamma_a - \Gamma_{C',HN})^{-1}$. Consequently, one

should choose T_C similar to typical T_2^* of MQ coherences present during CT period in order to minimize the absolute error of CCR rate. For the investigated sample of cBASP1, this was estimated to $\approx$ 90 ms by performing a series of 1D spectra with a variable T_C. Noteworthy is that without NUS, there is no flexibility of T_C, or it is only partially utilized for C' evolution because of sampling limitations.

The values of $\Gamma_{C',HN}$ determined for consequtive a. a. residues for both samples are shown in Fig. 4.39a,b. It should be emphasised that suppression of NUS artefacts has lead to

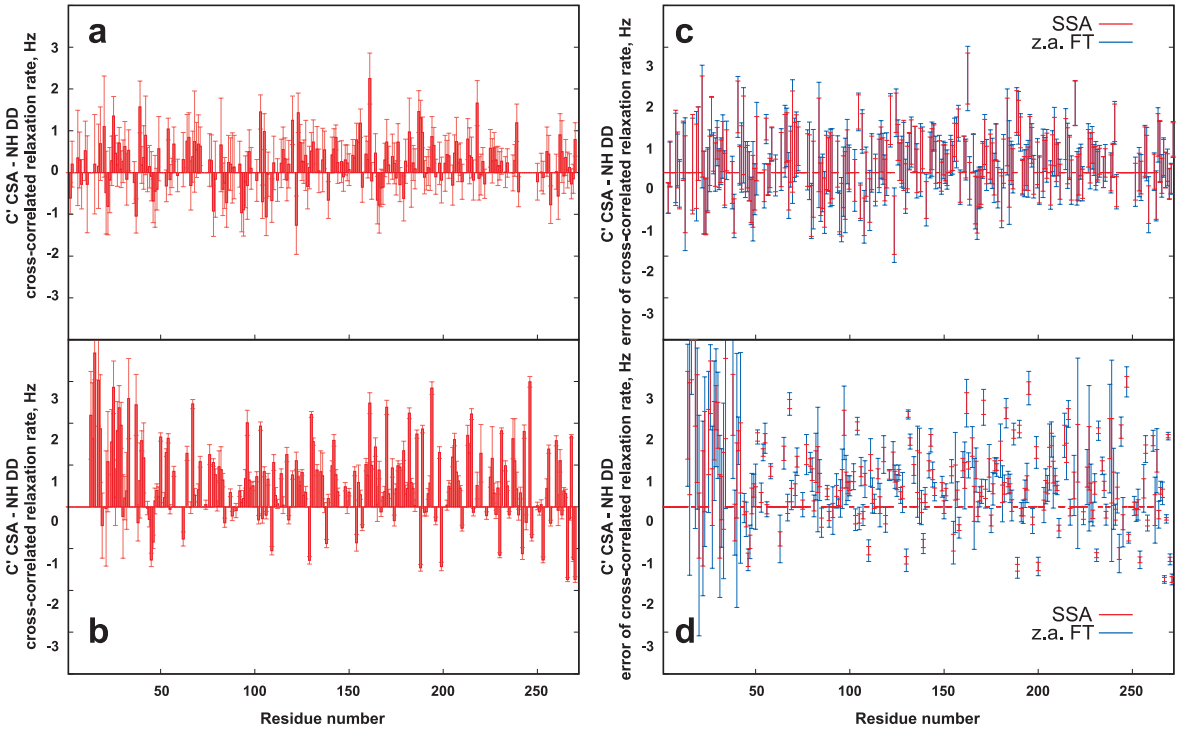


Fig. 4.39 The CCR (CSA C' - DD NH) rates determined for individual a. a. of BASP1 at pH=2 (a) and pH=6 (b). Panels (c, d) show the errors of CCR rates in spectra obtained with SSA (*red*) and zero-augmented FT (*blue*). The importance of artefact suppression for the sample at pH=6 is emphasised.

increase of average S/N ratio of relevant peaks by a factor of 2.7 and 1.3 for BASP1 at pH=6 and 2, respectively. According to the eq. 4.16, the errors of CCR rates were decreased by the same factor (Fig. 4.39c,d), which greatly enhanced the precision of obtained CCR rates. The suppression of artefacts was virtually complete, and different gains in S/N reflect the intrinsic sensitivity associated with various sample concentrations.

It is quite interesting, that despite rather significant relative errors of single CCR rates and the incomplete data, it is easy to detect a substantial differences between results for BASP1 at pH=2 and 6. It can be concluded from Fig. 4.39a,b, and is even more evident from superposition of normalized histograms of CCR rates in Fig. 4.40 that increase of pH induces systematic **positive shift**, and broadening of CCR distributions. This has been interpreted as follows: at low pH BASP1 tends to sample type-II β-turns and α-helices (alternating signs of CCR) while increasing of pH enhances the populations of type-I β-turns as well as polyprolin-II helices. This is in agreement with independent DOSY and DEER data that confirm quite unexpected overall compaction of the BASP1 structural ensemble at low pH (Geist et al 2013a). The dramatic change in conformational preferences can be associated with significant fraction of acidic residues: Glu (16%) and Asp (6%), and also with high proline content (9%).

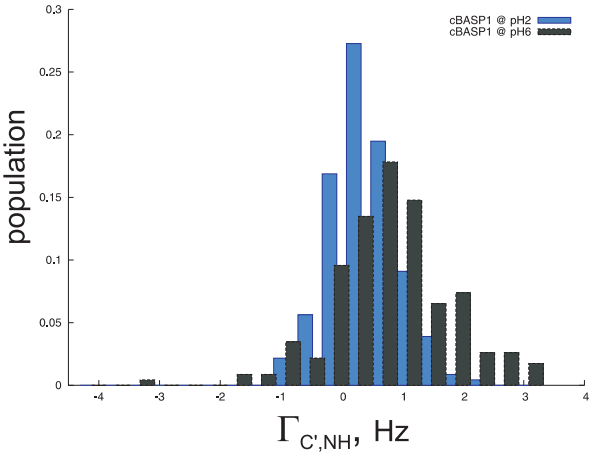


Fig. 4.40 The normalized histograms of ca. 220 CCR rates obtained for BASP1. Superimposed are distributions for the sample at pH=2 (*blue*) and pH=6 (*black*).

To sum up, it was demonstrated that CCR (C', NH) can reveal pH-induced shifts of β-turn populations in IDPs, which could not be concluded from currently accessible NMR data (e.g., chemical shifts, PREs, ^{15}N-NOEs or R_2 (^{15}N) relaxation rates). The successful determination of CCR values for challenging BASP1 required the joint use of recently developed approaches: increased dimensionality, NUS and SSA. It is also anticipated that other kinds of CCR rates will be accessible for IDPs as soon as appropriate 4D extensions of known pulse sequences are proposed. This could enable utilisation of multiple CCR values for studying conformational distributions of IDPs, similarly to the approach proposed for stably folded proteins (Kloiber et al 2002). The novelty of the presented approach was appreciated by the publication of the above results in the prestigious chemistry journal - *Angewandte Chemie - International Edition* (Stanek et al 2013c).

Bibliography

Akbey U, Lange S, Trent Franks W, Linser R, Rehbein K, Diehl A, van Rossum BJ, Reif B, Oschkinat H (2010) Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy. J Biomol NMR 46(1):67–73

Amero C, Asunción Durá M, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin M, Vernet T, Franzetti B, Boisbouvier J (2011) A systematic mutagenesis-driven strategy for site-resolved nmr studies of supramolecular assemblies. J Biomol NMR 50(3):229–236

Ardenkjær-Larsen JH, Fridlund B, Gram A, Hansson G, Hansson L, Lerche MH, Servin R, Thaning M, Golman K (2003) Increase in signal-to-noise ratio of $>$ 10,000 times in liquid-state NMR. P Natl Acad Sci USA 100(18):10,158–10,163

Atreya HS, Szyperski T (2004) G-matrix fourier transform NMR spectroscopy for complete protein resonance assignment. P Natl Acad Sci USA 101(26):9642–9647

Augustyniak R, Balayssac S, Ferrage F, Bodenhausen G, Lequin O (2011) ^{1}H, ^{13}C and ^{15}N resonance assignment of a 114-residue fragment of Engrailed 2 homeoprotein, a partially disordered protein. Biomol NMR Assigm 5(2):229–231

Augustyniak W (2012) Structure determination using four-dimensional non-uniformly sampled NOESY spectra, UW internal report

Augustyniak W, Brzezinska AA, Pijning T, Wienk H, Boelens R, Dijkstra BW, Reetz MT (2012) Biophysical characterization of mutants of *Bacillus subtilis* lipase evolved for thermostability: Factors contributing to increased activity retention. Protein Sci 21(4):487–497

Augustyniak W, Wienk H, Boelens R, Reetz MT (2013) ^{1}H, ^{13}C and ^{15}N resonance assignments of wild-type *Bacillus subtilis* Lipase A and its mutant evolved towards thermostability. Biomol NMR Assigm 7(2):249–252

Ayala I, Sounier R, Usé N, Gans P, Boisbouvier J (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J Biomol NMR 43(2):111–119

Ayala I, Hamelin O, Amero C, Pessey O, Plevin MJ, Gans P, Boisbouvier J (2012) An optimized isotopic labelling strategy of isoleucine-γ_2 methyl groups for solution NMR studies of high molecular weight proteins. Chem Commun 48:1434–1436

Barna J, Tan S, Laue E (1988) Use of clean in conjunction with selective data sampling for 2D NMR experiment. J Magn Reson 78:327–332

Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. J Magn Reson 73(1):69–77

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. Biochemistry-US 39(18):5355–5365

Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE (1990a) Practical aspects of proton carbon carbon proton 3-Dimensional correlation spectroscopy of ^{13}C-labeled proteins. J Magn Reson 87(3):620–627

Bax A, Clore GM, Gronenborn AM (1990b) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. J Magn Reson 88(2):425–431

Bermel W, Bertini I, Felli IC, Lee YM, Luchinat C, Pierattelli R (2006) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128(12):3918–3919

Bermel W, Bertini I, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, Stanek J (2012) Speeding up

sequence specific assignment of IDPs. J Biomol NMR 53(4):293–301

Bertini I, Duma L, Felli IC, Fey M, Luchinat C, Pierattelli R, Vasos PR (2004) A heteronuclear direct-detection NMR spectroscopy experiment for protein-backbone assignment. Angew Chem Int Edit 43(17):2257–2259

Bertini I, Emsley L, Lelli M, Luchinat C, Mao J, Pintacuda G (2010) Ultrafast MAS solid-state NMR permits extensive ^{13}C and ^{1}H detection in paramagnetic metalloproteins. J Am Chem Soc 132(16):5558–5559

Bertini I, Luchinat C, Parigi G, Ravera E, Reif B, Turano P (2011) Solid-state NMR of proteins sedimented by ultracentrifugation. P Natl Acad Sci USA 108(26):10,396–10,399

Blackford LS, Demmel J, Dongarra J, Duff I, Hammarling S, Henry G, Heroux M, Kaufman L, Lumsdaine A, Petitet A, Pozo R, Remington K, Whaley RC (2002) An Updated Set of Basic Linear Algebra Subprograms (BLAS). ACM T Math Software 28:135–151

Bostock MJ, Holland DJ, Nietlispach D (2012) Compressed sensing reconstruction of undersampled 3D NOESY spectra: application to large membrane proteins. J Biomol NMR 54(1):15–32

Bracewell RN (2000) The Fourier Transform and Its Applications, 3rd edn. McGraw-Hill, Singapore

Bretthorst GL (1990) Bayesian-analysis 3. Applications to NMR signal-detection, model selection, and parameter-estimation. J Magn Reson 88(3):571–595

Bro R (1997) PARAFAC. Tutorial and applications. Chemometr Intell Lab 38(2):149 – 171

Brutscher B, Boisbouvier J, Pardi A, Marion D, Simorre JP (1998) Improved sensitivity and resolution in ^{1}H–^{13}C NMR experiments of RNA. J Am Chem Soc 120(46):11,845–11,851

Candès E, Wakin M (2008) An introduction to compressive sampling. IEEE Signal Proc Mag 25(2):21–30

Candès E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE T Inform Theory 52(2):489–509

Candès E, Wakin M, Boyd S (2008) Enhancing sparsity by reweighted l_1 minimization. J Fourier Anal Appl 14(5-6):877–905

Caroll JD, Pruzansky S (1984) Research methods for multimode data analysis. Praeger, New York, NY, pp 372–402

Carroll J, Chang JJ (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. Psychometrika 35(3):283–319

Cavanagh J, Fairbrother WJ, Palmer III AG, Skelton NJ, Rance M (2007) Protein NMR Spectroscopy, 2nd edn. Elsevier Academic Press

Cevec M, Thibaudeau C, Plavec J (2010) NMR structure of the let-7 miRNA interacting with the site LCS1 of lin-41 mRNA from *Caenorhabditis elegans*. Nucleic Acids Res 38(21):7814–7821

Chartrand R, Staneva V (2008) Restricted isometry properties and nonconvex compressive sensing. Inverse Probl 24(3):035,020

Chartrand R, Yin W (2008) Iteratively reweighted algorithms for compressive sensing. In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp 3869–3872

Chevelkov V, Rehbein K, Diehl A, Reif B (2006) Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. Angew Chem Int Edit 45(23):3878–3881

Chiarparin E, Pelupessy P, Ghose R, Bodenhausen G (2000) Relative orientation of $C^{\alpha}H^{\alpha}$-bond vectors of successive residues in proteins through cross-correlated relaxation in NMR. J Am Chem Soc 122(8):1758–1761

Chylla RA, Markley JL (1993) Improved frequency resolution in multidimensional constant-time experiments by multidimensional Bayesian-analysis. J Biomol NMR 3(5):515–533

Chylla RA, Markley JL (1995) Theory and application of the maximum-likelihood principle to NMR parameter-estimation of multidimensional NMR data. J Biomol NMR 5(3):245–258

Chylla RA, Volkman BF, Markley JL (1998) Practical model fitting approaches to the direct extraction of NMR parameters simultaneously from all dimensions of multidimensional NMR spectra. J Biomol NMR 12(2):277–297

Clore GM, Kay LE, Bax A, Gronenborn AM (1991) Four-dimensional ^{13}C,^{13}C-edited Nuclear Overhauser Enhancement spectroscopy of a protein in solution: Application to Interleukin 1-β. Biochemistry-US 30(1):12–18

Clowes RT, Boucher W, Hardman CH, Domaille PJ, Laue ED (1993) A 4D HCC(CO)NNH experiment for

the correlation of aliphatic side-chain and backbone resonances in $^{13}C/^{15}N$ labeled proteins. J Biomol NMR 3(3):349–354

Coggins BE, Zhou P (2007) Sampling of the NMR time domain along concentric rings. J Magn Reson 184(2):207 – 221

Coggins BE, Zhou P (2008) High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. J Biomol NMR 42(4):225–39

Coggins BE, Venters RA, Zhou P (2005) Filtered backprojection for the reconstruction of a high-resolution (4,2)D CH_3–NH NOESY spectrum on a 29 kDa protein. J Am Chem Soc 127(33):11,562–11,563

Coggins BE, Werner-Allen JW, Yan A, Zhou P (2012) Rapid protein global fold determination using ultrasparse sampling, high-dynamic range artifact suppression, and time-shared NOESY. J Am Chem Soc 134(45):18,619–18,630

Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex fourier series. Math Comput 19:297–901

Dalvit C (1992) ^{1}H to ^{15}N polarization transfer via ^{1}H chemical-shift anisotropy – ^{1}H-^{15}N dipole-dipole cross correlation. J Magn Reson (1969) 97(3):645 – 650

Daniell G, Hore P (1989) Maximum entropy and NMR - a new approach. J Magn Reson 84:515–536

Davies SJ, Bauer C, Hore PJ, Freeman R (1988) Resolution enhancement by nonlinear data-processing - HOGWASH and the maximum-entropy method. J Magn Reson 76(3):476–493

De Paëpe G, Giraud N, Lesage A, Hodgkinson P, Böckmann A, Emsley L (2003) Transverse dephasing optimized solid-state NMR spectroscopy. J Am Chem Soc 125(46):13,938–13,939

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–93

Diercks T, Orekhov V (2005) qTROSY – a novel scheme for recovery of the anti-TROSY magnetisation. J Biomol NMR 32(2):113–127

Diercks T, Truffault V, Coles M, Millett O (2010) Diagonal-free 3D/4D HN,HN-TROSY-NOESY-TROSY. J Am Chem Soc 132(7):2138–2139

Ding K, Gronenborn AM (2004) Protein backbone $^{1}H^{N}$-$^{13}C^{\alpha}$ and ^{15}N–$^{13}C^{\alpha}$ residual dipolar and J couplings: New constraints for NMR structure determination. J Am Chem Soc 126(20):6232–6233

Donoho D (2006) Compressed sensing. IEEE T Inform Theory 52(4):1289–1306

Donoho D, Johnstone I, Stern A, Hoch J (1990) Does the maximum entropy method improve sensitivity? Proc Natl Acad Sci USA 87:5066–5068

Driscoll PC, Gronenborn AM, Clore GM (1989) The influence of stereospecific assignments on the determination of 3-Dimensional structures of proteins by nuclear magnetic-resonance spectroscopy – application to the sea-anemone protein Bds-I. FEBS Lett 243(2):223–233

Drori I (2007) Fast minimization by iterative thresholding for multidimensional NMR spectroscopy. Eurasip J Adv Sig Pr 2007(1):020,248

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97(1):135–148

Ernst RR, Bodenhausen G, Alexander W (1987) Principles of nuclear magnetic resonance in one and two dimensions. Oxford Univ. Press, London/New York

Farjon J, Boisbouvier J, Schanda P, Pardi A, Simorre JP, Brutscher B (2009) Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. J Am Chem Soc 131(24):8571–8577

Farmer I BennettT, Mueller L (1994) Simultaneous acquisition of [^{13}C,^{15}N]- and [^{15}N,^{15}N]-separated 4D gradient-enhanced NOESY spectra in proteins. J Biomol NMR 4(5):673–687

Favier A, Brutscher B (2011) Recovering lost magnetization: Polarization enhancement in biomolecular NMR. J Biomol NMR 49(1):9–15

Fiala R, Jiang F, Sklenář V (1998) Sensitivity optimized HCN and HCNCH experiments for $^{13}C/^{15}N$ labeled oligonucleotides. J Biomol NMR 12(3):373–383

Fischer M, Kloiber K, Häusler J, Ledolter K, Konrat R, Schmid W (2007) Synthesis of a ^{13}C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. ChemBioChem 8(6):610–612

Folmer RHA, Hilbers CW, Konings RNH, Nilges M (1997) Floating stereospecific assignment revisited: Application to an 18 kDa protein and comparison with J-coupling data. J Biomol NMR 9(3):245–258

Frieden BR (1972) Restoring with maximum likelihood and maximum entropy. J Opt Soc Am 62(4):511–518

Frigo M, Johnson SG (2005) The design and implementation of FFTW3. P IEEE 93(2):216–231

Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) NMR spectroscopy of RNA. ChemBioChem 4(10):936–962

Frueh D, Vosburg D, Walsh C, Wagner G (2006a) Determination of all NOEs in ^{1}H–^{13}C–Me-ILV-U–^{2}H–^{15}N proteins with two time-shared experiments. J Biomol NMR 34(1):31–40

Frueh DP, Sun ZYJ, Vosburg DA, Walsh CT, Hoch JC, Wagner G (2006b) Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. J Am Chem Soc 128(17):5757–5763

Frydman L, Scherf T, Lupulescu A (2002) The acquisition of multidimensional NMR spectra within a single scan. P Natl Acad Sci USA 99(25):15,858–62

Frydman L, Lupulescu A, Scherf T (2003) Principles and features of single-scan two-dimensional NMR spectroscopy. J Am Chem Soc 125(30):9204–9217

Gal M, Mishkovsky M, Frydman L (2006) Real-time monitoring of chemical transformations by ultrafast 2D NMR spectroscopy. J Am Chem Soc 128(3):951–956

Gal M, Schanda P, Brutscher B, Frydman L (2007) UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. J Am Chem Soc 129(5):1372–1377

Gans P, Hamelin O, Sounier R, Ayala I, Durá M, Amero C, Noirclerc-Savoye M, Franzetti B, Plevin M, Boisbouvier J (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem Int Edit 49(11):1958–1962

Gardiennet C, Schütz AK, Hunkeler A, Kunert B, Terradot L, Böckmann A, Meier BH (2012) A sedimented sample of a 59 kDa dodecameric helicase yields high-resolution solid-state NMR spectra. Angew Chem Int Edit 51(31):7855–7858

Gardner KH, Kay LE (1997) Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ_1 Methyl) isoleucine into proteins for multidimensional NMR studies. J Am Chem Soc 119(32):7599–7600

Gardner KH, Kay LE (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Bioph Biom 27:357–406

Geist L, Henen MA, Haiderer S, Schwarz TC, Kurzbach D, Zawadzka-Kazimierczuk A, Saxena S, Żerko S, Koźminski W, Hinderberger D, Konrat R (2013a) Protonation-dependent conformational variability of intrinsically disordered proteins. Protein Sci 22(9):1196–1205

Geist L, Zawadzka-Kazimierczuk A, Saxena S, Żerko S, Koźminski W, Konrat R (2013b) ^{1}H, ^{13}C and ^{15}N resonance assignments of human BASP1. Biomol NMR Assigm 7(2):315–319

Gerchberg R (1974) Super-resolution through error energy reduction. Opt Acta 21(9):709–720

Giraudeau P, Shrot Y, Frydman L (2009) Multiple ultrafast, broadband 2D NMR spectra of hyperpolarized natural products. J Am Chem Soc 131(39):13,902–13,903

Güntert P (2004) Automated NMR structure calculation with CYANA. In: Downing A (ed) Protein NMR Techniques, Method Mol Cell Bio, vol 278, Humana Press, pp 353–378

Güntert P, Braun W, Billeter M, Wüthrich K (1989) Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein-structure determinations in solution. J Am Chem Soc 111(11):3997–4004

Goddart TD, Kneller DG (1989-2008) SPARKY 3, University of California, San Francisco

Goldman M (1984) Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. J Magn Reson (1969) 60(3):437 – 452

Goto N, Gardner K, Mueller G, Willis R, Kay L (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ_1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins. J Biomol NMR 13(4):369–374

Griesinger C, Bennati M, Vieth H, Luchinat C, Parigi G, Höfer P, Engelke F, Glaser S, Denysenkov V, Prisner T (2012) Dynamic nuclear polarization at high magnetic fields in liquids. Prog Nucl Mag Res Sp 64(0):4 – 28

Griffey RH, Redfield AG (1987) Proton-detected heteronuclear edited and correlated nuclear magnetic resonance and nuclear Overhauser effect in solution. Q Rev Biophys 19:51–82

Griffin RG (2010) Spectroscopy: Clear signals from surfaces. Nature 468(7322):381–382

Grzesiek S, Bax A (1995) Spin-locked multiple quantum coherence for signal enhancement in heteronuclear multidimensional NMR experiments. J Biomol NMR 6(3):335–339

Grzesiek S, Anglister J, Bax A (1993) Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization. J Magn Reson Ser B 101(1):114 – 119

Grzesiek S, Kuboniwa H, Hinck AP, Bax A (1995) Multiple-quantum line narrowing for measurement of H^{α}-H^{β} J-couplings in isotopically enriched proteins. J Am Chem Soc 117(19):5312–5315

Gschwind R, Gemmecker G, Kessler H (1998) A spin system labeled and highly resolved ed-H(CCO)NH-TOCSY experiment for the facilitated assignment of proton side chains in partially deuterated samples. J Biomol NMR 11(2):191–198

Gull SF, Daniell GJ (1978) Image-reconstruction from incomplete and noisy data. Nature 272(5655):686–690

Gutmanas A, Jarvoll P, Orekhov V, Billeter M (2002) Three-way decomposition of a complete 3D ^{15}N-NOESY-HSQC. J Biomol NMR 24(3):191–201

Harshmann R (1970) Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis. UCLA Working Papers in Phonetics 16:1–84

Harshmann R, Lundy M (1984) Data preprocessing and the extended PARAFAC model. In: Law H, Snyder CW, Hattie J, McDonald R (eds) Research Methods for Multimode Data Analysis, Praeger, New York

Hausser K, Stehlik D (1968) Dynamic nuclear polarization in liquids. In: WAUGH JS (ed) Advances in Magnetic Resonance, Advances in Magnetic and Optical Resonance, vol 3, Academic Press, pp 79 – 139

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319(1):209–227

Högbom J (1974) Aperture synthesis with a non-regular distribution of interferometer baselines. Astron Astrophys Suppl Ser 15:417–426

Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G (2008a) Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. Science 321(5893):1206–1210

Hiller S, Joss R, Wider G (2008b) Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY). J Am Chem Soc 130(36):12,073–12,079

Hiller S, Ibraghimov I, Wagner G, Orekhov VY (2009) Coupled decomposition of four-dimensional NOESY spectra. J Am Chem Soc 131(36):12,970–12,978

Hoch J, Stern A (1996) NMR Data Processing. Wiley-Interscience, New York

Hoch J, Stern A, Donoho D, Johnstone I (1990) Maximum entropy reconstruction of complex (phase-sensitive) spectra. J Magn Reson 86:236–246

Hoch JC (1985) Maximum-entropy signal-processing of two-dimensional NMR data. J Magn Reson 64(3):436–440

Hoch JC, Stern AS (2001) Maximum entropy reconstruction, spectrum analysis and deconvolution in multidimensional nuclear magnetic resonance. Method Enzymol 338:159–178

Hoch JC, Maciejewski MW, Filipovic B (2008) Randomization improves sparse sampling in multidimensional NMR. J Magn Reson 193(2):317 – 320

Holland DJ, Bostock MJ, Gladden LF, Nietlispach D (2011) Fast multidimensional NMR spectroscopy using compressed sensing. Angew Chem Int Edit 50(29):6548–6551

Hologne M, Chevelkov V, Reif B (2006) Deuterated peptides and proteins in MAS solid-state NMR. Prog Nucl Mag Res Sp 48(4):211 – 232

Hore P (1985) NMR data processing using the maximum entropy method. J Magn Reson 62:561–567

Hu KN, Yu Hh, Swager TM, Griffin RG (2004) Dynamic nuclear polarization with biradicals. J Am Chem Soc 126(35):10,844–10,845

Hu W, Jiang L (1999) Multiple-quantum HCN-CCH-TOCSY experiment for $^{13}C/^{15}N$ labeled RNA oligonucleotides. J Biomol NMR 15(4):289–293

Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luithardt H, Fejzo J, Chorev M, Aktas H, Colson K, Falchuk KH, Halperin JA, Wagner G (2007) Ultrahigh-resolution ^{1}H-^{13}C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. J Am Chem Soc 129(16):5108–5116

Hyberts SG, Frueh DP, Arthanari H, Wagner G (2009) FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. J Biomol NMR 45(3):283–294

Hyberts SG, Takeuchi K, Wagner G (2010) Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. J Am Chem Soc 132(7):2145–2147

Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. J Biomol NMR 52(4):315–327

Hyberts SG, Robson SA, Wagner G (2013) Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J Biomol NMR 55(2):167–178

Ibraghimov I (2002) Application of the three-way decomposition for matrix compression. Numer Linear Algebr 9(6-7):551–565

Ibraghimov IV (1999) A new approach to solution of general singular vector decomposition problem. In: Tyrtyshnikov E (ed) Matrix Methods and Algorithms, Inst. Comput. Mathematics RAS, Moscow, pp 194–202

Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry-US 29(19):4659–4667

Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P, Matthews S (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. J Am Chem Soc 129(50):15,428–15,429

Jaravine V, Ibraghimov I, Orekhov VY (2006) Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. Nat Methods 3(8):605–607

Jaravine VA, Orekhov VY (2006) Targeted acquisition for real-time NMR spectroscopy. J Am Chem Soc 128(41):13,421–13,426

Jaravine VA, Zhuravleva AV, Permi P, Ibraghimov I, Orekhov VY (2008) Hyperdimensional NMR spectroscopy with nonlinear sampling. J Am Chem Soc 130(12):3927–3936

Jeannerat D (2003) High resolution in heteronuclear ^{1}H–^{13}C nmr experiments by optimizing spectral aliasing with one-dimensional carbon data. Magn Reson Chem 41(1):3–17

Jeannerat D (2007) Computer optimized spectral aliasing in the indirect dimension of ^{1}H–^{13}C heteronuclear 2D NMR experiments. A new algorithm and examples of applications to small molecules. J Magn Reson 186(1):112 – 122

Jee J, Güntert P (2003) Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. J Struct Funct Genomics 4(2-3):179–189

Jiang L, Coggins BE, Zhou P (2005) Rapid assignment of protein side chain resonances using projection–reconstruction of (4,3)D HC(CCO)NH and intra-HC(C)NH experiments. J Magn Reson 175(1):170 – 176

Jordan JB, Kovacs H, Wang Y, Mobli M, Luo R, Anklin C, Hoch JC, Kriwacki RW (2006) Three-dimensional ^{13}C-detected CH_3-TOCSY using selectively protonated proteins: Facile methyl resonance assignment and protein structure determination. J Am Chem Soc 128(28):9119–9128

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440(7080):52–57

Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc 85(18):2870–2871

Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114(26):10,663–10,665

Kay LE, Marion D, Bax A (1989) Practical aspects of 3D heteronuclear NMR of proteins. J Magn Reson 84(1):72–84

Kay LE, Clore GM, Bax A, Gronenborn AM (1990a) Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1β in solution. Science 249(4967):411–414

Kay LE, Ikura M, Tschudin R, Bax A (1990b) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson (1969) 89(3):496 – 514

Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. J Magn Reson Ser

B 101(3):333–337

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Edit 50(24):5556–5559

Kazimierczuk K, Orekhov VY (2012) A comparison of convex and non-convex compressed sensing applied to multidimensional NMR. J Magn Reson 223(0):1–10

Kazimierczuk K, Koźmiński W, Zhukov I (2006a) Two-dimensional Fourier transform of arbitrarily sampled NMR data sets. J Magn Reson 179(2):323 – 328

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006b) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36(3):157–68

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2007) Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets. J Magn Reson 188(2):344–56

Kazimierczuk K, Zawadzka A, Koźmiński W (2008a) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192(1):123–30

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2008b) Determination of spin-spin couplings from ultrahigh resolution 3D NMR spectra obtained by optimized random sampling and multidimensional Fourier transformation. J Am Chem Soc 130(16):5404–5

Kazimierczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197(2):219–28

Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W (2010) Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J Magn Reson 205(2):286–92

Kazimierczuk K, Misiak M, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2012) Generalized Fourier transform for non-uniform sampled data. Top Curr Chem 316:79–124

Keeler J (1984) Elimination of truncation artifacts from NMR-spectra - application to C-13 multiplicity determination by two-dimensional spectroscopy. J Magn Reson 56(3):463–470

Keeler J (2005) Understanding NMR spectroscopy. Wiley, Chichester, UK

Kloiber K, Konrat R (2000a) Measurement of the protein backbone dihedral angle φ based on quantification of remote CSA/DD interference in inter-residue ^{13}C'(i-1)-^{13}C$^{\alpha}$ (i) multiple-quantum coherences. J Biomol NMR 17(3):265–268

Kloiber K, Konrat R (2000b) Peptide plane torsion angles in proteins through intraresidue ^{1}H-^{15}N-^{13}C' dipole-CSA relaxation interference: Facile discrimination between type-I and type-II β-turns. J Am Chem Soc 122(48):12,033–12,034

Kloiber K, Schüler W, Konrat R (2002) Automated NMR determination of protein backbone dihedral angles from cross-correlated spin relaxation. J Biomol NMR 22(4):349–363

Knight MJ, Webber AL, Pell AJ, Guerry P, Barbet-Massin E, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Emsley L, Lesage A, Herrmann T, Pintacuda G (2011) Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state MAS NMR spectroscopy. Angew Chem Int Edit 50(49):11,697–11,701

Knight MJ, Pell AJ, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Herrmann T, Emsley L, Pintacuda G (2012) Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. P Natl Acad Sci USA 109(28):11,095–11,100

Koehl P (1999) Linear prediction spectral analysis of NMR data. Prog Nucl Mag Res Sp 34(3–4):257 – 299

Korzhnev D, Ibraghimov I, Billeter M, Orekhov V (2001) MUNIN: Application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. J Biomol NMR 21(3):263–268

Kosiński K (2013) Implementation of the signal separation algorithm for five-dimensional NMR spectra. Master's thesis, University of Warsaw, Faculty of Chemistry

Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes – a leap in NMR technology. Prog Nucl Mag Res Sp 46(2–3):131 – 155

Kroonenberg P, Leeuw J (1980) Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika 45(1):69–97

Kruskal JB (1977) Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl 18(2):95 – 138

Kumar A, Brown SC, Donlan ME, Meier BU, Jeffs PW (1991) Optimization of two-dimensional NMR by matched accumulation. J Magn Reson (1969) 95(1):1–9

Kupče E, Freeman R (2008) Molecular structure from a single NMR experiment. J Am Chem Soc 130(32):10,788–10,792

Kupče E, Kay LE, Freeman R (2010) Detecting the "afterglow" of ^{13}C NMR in proteins using multiple receivers. J Am Chem Soc 132(51):18,008–18,011

Kupče r, Freeman R (2005) Fast multidimensional NMR: radial sampling of evolution space. J Magn Reson 173(2):317–21

Kupče r, Kay LE (2012) Parallel acquisition of multi-dimensional spectra in protein NMR. J Biomol NMR 54(1):1–7

Kupče r, Freeman R, John BK (2006) Parallel acquisition of two-dimensional NMR spectra of several nuclear species. J Am Chem Soc 128(30):9606–9607

Kızılsavaş G, Saxena S, Żerko S, Koźmiński W, Bister K, Konrat R (2013) ^{1}H, ^{13}C, and ^{15}N backbone and side chain resonance assignments of the C-terminal DNA binding and dimerization domain of v-Myc. Biomol NMR Assigm 7(2):321–324

Laage S, Sachleben JR, Steuernagel S, Pierattelli R, Pintacuda G, Emsley L (2009) Fast acquisition of multi-dimensional spectra in solid-state NMR enabled by ultra-fast MAS. J Magn Reson 196(2):133 – 141

Laue E, Mayger M, Skilling J, Staunton J (1986) Reconstruction of phase sensitive 2D NMR spectra by maximum entropy. J Magn Reson 68:14–28

Laue ED, Skilling J, Staunton J (1985a) Maximum-entropy reconstruction of spectra containing antiphase peaks. J Magn Reson 63(2):418–424

Laue ED, Skilling J, Staunton J, Sibisi S, Brereton RG (1985b) Maximum entropy method in nuclear magnetic resonance spectroscopy. J Magn Reson 62:437–452

Lee MK, Gal M, Frydman L, Varani G (2010) Real-time multidimensional NMR follows RNA folding with second resolution. P Natl Acad Sci USA 107(20):9192–9197

Lemaster DM (1987) Chiral-beta and random fractional deuteration for the determination of protein side-chain conformation by NMR. FEBS Lett 223(1):191–196

Lemaster DM (1990) Deuterium labeling in NMR structural-analysis of larger proteins. Q Rev Biophys 23(2):133–174

LeMaster DM (1994) Isotope labeling in solution protein assignment and structural analysis. Prog Nucl Mag Res Sp 26, Part 4(0):371 – 419

LeMaster DM, Kushlan DM (1996) Dynamical mapping of E. *coli* thioredoxin via ^{13}C NMR relaxation analysis. J Am Chem Soc 118(39):9255–9264

Lesage A, Bardet M, Emsley L (1999) Through-bond carbon–carbon connectivities in disordered solids by NMR. J Am Chem Soc 121(47):10,987–10,993

Lesage A, Lelli M, Gajan D, Caporini MA, Vitzthum V, Miéville P, Alauzun J, Roussey A, Thieuleux C, Mehdi A, Bodenhausen G, Coperet C, Emsley L (2010) Surface enhanced NMR spectroscopy by dynamic nuclear polarization. J Am Chem Soc 132(44):15,459–15,461

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187(1):163–169

Lescop E, Kern T, Brutscher B (2010) Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: Example of longitudinal-relaxation-enhanced BEST-type ^{1}H-^{15}N correlation experiments. J Magn Reson 203(1):190–198

Löhr F, Reckel S, Karbyshev M, Connolly P, Abdul-Manan N, Bernhard F, Moore J, Dötsch V (2012) Combinatorial triple-selective labeling as a tool to assist membrane protein backbone resonance assignment. J Biomol NMR 52(3):197–210

Li KB, Stern AS, Hoch JC (1998) Distributed parallel processing for multidimensional maximum entropy reconstruction. J Magn Reson 134(1):161 – 163

Liang B, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. J Am Chem Soc 128(13):4389–4397

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104(17):4546–4559

Liu W, Zheng Y, Cistola D, Yang D (2003) Measurement of methyl ^{13}C-^{1}H cross-correlation in uniformly ^{13}C-, ^{15}N-, labeled proteins. J Biomol NMR 27(4):351–364

Logan TM, Olejniczak ET, Xu RX, Fesik SW (1992) Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. FEBS Lett 314(3):413 – 418

Lozhko D, Stanek J, Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźminski W, Zhukov I, Kornelyuk A (2013) ^{1}H, ^{13}C, and ^{15}N chemical shifts assignments for human endothelial monocyte-activating polypeptide EMAP II. Biomol NMR Assigm 7:25–29

Luan T, Jaravine V, Yee A, Arrowsmith CH, Orekhov VY (2005) Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. J Biomol NMR 33(1):1–14

Maciejewski MW, Qui HZ, Rujan I, Mobli M, Hoch JC (2009) Nonuniform sampling and spectral aliasing. J Magn Reson 199(1):88 – 93

Maly T, Debelouchina GT, Bajaj VS, Hu KN, Joo CG, Mak-Jurkauskas ML, Sirigiri JR, van der Wel PCA, Herzfeld J, Temkin RJ, Griffin RG (2008) Dynamic nuclear polarization at high magnetic fields. J Chem Phys 128(5)

Marchetti A, Jehle S, Felletti M, Knight MJ, Wang Y, Xu ZQ, Park AY, Otting G, Lesage A, Emsley L, Dixon NE, Pintacuda G (2012) Backbone assignment of fully protonated solid proteins by ^{1}H detection and ultrafast magic-angle-spinning NMR spectroscopy. Angew Chem Int Edit 51(43):10,756–10,759

Marino JP, Diener JL, Moore PB, Griesinger C (1997) Multiple-quantum coherence dramatically enhances the sensitivity of CH and CH_2 correlations in uniformly ^{13}C-labeled RNA. J Am Chem Soc 119(31):7361–7366

Markus M, Dayie K, Matsudaira P, Wagner G (1994) Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on Villin 14T. J Magn Reson Ser B 105(2):192 – 195

Matsuki Y, Eddy MT, Herzfeld J (2009) Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra. J Am Chem Soc 131(13):4648–4656

Matsuki Y, Eddy MT, Griffin RG, Herzfeld J (2010) Rapid three-dimensional MAS NMR spectroscopy at critical sensitivity. Angew Chem Int Edit 49(48):9215–9218

Mayzel M, Rosenlöw J, Isaksson L, Orekhov V (2014) Time-resolved multidimensional NMR with non-uniform sampling. J Biomol NMR 58(2):129–139

McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear-magnetic-resonance spectra of proteins. Q Rev Biophys 23(1):1–38

Meissner A, Sørensen OW (1999) The role of coherence transfer efficiency in design of TROSY-type multidimensional NMR experiments. J Magn Reson 139(2):439 – 442

Meissner A, Schulte-Herbrüggen T, Briand J, Sørensen OW (1998) Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times. Mol Phys 95(6):1137–1142

Mobli M, Stern AS, Hoch JC (2006) Spectral reconstruction methods in fast NMR: Reduced dimensionality, random sampling and maximum entropy. J Magn Reson 182(1):96 – 105

Mobli M, Maciejewski M, Gryk M, Hoch J (2007) Automatic maximum entropy spectral reconstruction in NMR. J Biomol NMR 39:133–139

Mobli M, Stern AS, Bermel W, King GF, Hoch JC (2010) A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. J Magn Reson 204(1):160–164

Montelione GT, Lyons BA, Emerson SD, Tashiro M (1992) An efficient triple resonance experiment using ^{13}C isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 114(27):10,974–10,975

Morris GA, Freeman R (1979) Enhancement of nuclear magnetic resonance signals by polarization transfer. J Am Chem Soc 101(3):760–762

Motáčková V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. J Biomol NMR 48(3):169–177

Muhandiram DR, Farrow NA, Xu GY, Smallcombe SH, Kay LE (1993) A gradient ^{13}C NOESY-HSQC ex-

periment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H_2O. J Magn Reson Ser B 102(3):317–321

Mumenthaler C, Güntert P, Braun W, Wüthrich K (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. J Biomol NMR 10(4):351–362

Nietlispach D (2005) Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. J Biomol NMR 31(2):161–166

Nietlispach D, Clowes RT, Broadhurst RW, Ito Y, Keeler J, Kelly M, Ashurst J, Oschkinat H, Domaille PJ, Laue ED (1996) An approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration. J Am Chem Soc 118(2):407–415

Norwood TJ (1992) Multiple-quantum NMR methods. Prog Nucl Mag Res Sp 24:295–375

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50(1):1–11

Nováček J, Haba N, Chill J, Žídek L, Sklenář V (2012) 4D non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53(2):139–148

Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56(4):291–301

Nowakowski M, Jaremko L, Jaremko M, Zhukov I, Belczyk A, Bierzyński A, Ejchart A (2011) Solution NMR structure and dynamics of human apo-S100A1 protein. J Struct Biol 174(2):391–399

Nudelman I, Akabayov SR, Scherf T, Anglister J (2011) Observation of intermolecular interactions in large protein complexes by 2D-double difference nuclear Overhauser enhancement spectroscopy: Application to the 44 kDa interferon–receptor complex. J Am Chem Soc 133(37):14,755–14,764

Olejniczak ET, Xu RX, Fesik SW (1992) A 4D HCCH-TOCSY experiment for assigning the side chain ^{1}H and ^{13}C resonances of proteins. J Biomol NMR 2(6):655–659

Orbán-Németh Z, Henen M, Geist L, Żerko S, Saxena S, Stanek J, Koźmiński W, Propst F, Konrat R (2013) Backbone and partial side chain assignment of the microtubule binding domain of the MAP1B light chain. Biomol NMR Assigm pp 1–5

Orekhov V, Ibraghimov I, Billeter M (2001) MUNIN: A new approach to multi-dimensional NMR spectra interpretation. J Biomol NMR 20(1):49–60

Orekhov V, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27(2):165–173

Otten R, Chu B, Krewulak KD, Vogel HJ, Mulder FAA (2010) Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. J Am Chem Soc 132(9):2952–2960

Overhauser AW (1953) Polarization of nuclei in metals. Phys Rev 92:411–415

Palmer III AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. J Magn Reson (1969) 93(1):151 – 170

Pannetier N, Houben K, Blanchard L, Marion D (2007) Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins. J Magn Reson 186(1):142 – 149

Papoulis A (1975) A new algorithm in spectral analysis and band-limited extrapolation. IEEE T Circuits Syst 22(9):735–742

Parker MJ, Aulton-Jones M, Hounslow AM, Craven CJ (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. J Am Chem Soc 126(16):5020–5021

Pascal S, Muhandiram D, Yamazaki T, Formankay J, Kay L (1994) Simultaneous acquisition of ^{15}N- and ^{13}C-edited NOE spectra of proteins dissolved in H_2O. J Magn Reson Ser B 103(2):197 – 201

Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G (1999) Simultaneous determination of ψ and φ angles in proteins from measurements of cross-correlated relaxation effects. J Biomol NMR 14(3):277–280

Pervushin K, Riek R, Wider G, Wütrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates and avenue to NMR structures of very large biological macromolecules in solution. P Natl Acad Sci USA 94(23):12,366–12,371

Pervushin K, Riek R, Wider G, Wüthrich K (1998a) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labeled proteins. J Am Chem Soc 120(25):6394–6400

Pervushin K, Wider G, Wüthrich K (1998b) Single transition-to-single transition polarization transfer (ST2-PT) in [^{15}N, ^{1}H]-TROSY. J Biomol NMR 12(2):345–348

Pervushin K, Vögeli B, Eletsky A (2002) Longitudinal ^{1}H relaxation optimization in TROSY NMR spectroscopy. J Am Chem Soc 124(43):12,898–12,902

van Pouderoyen G, Eggert T, Jaeger K, Dijkstra B (2001) The crystal structure of B*acillus subtilis* lipase: a minimal α/β hydrolase fold enzyme. J Mol Biol 309:216–226

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical Recipies, The Art of Scientific Computing, 3rd edn. Cambridge University Press, New York, NY

Reddy J, Hosur R (2013) Parallel acquisition of 3D-HA(CA)NH and 3D-HACACO spectra. J Biomol NMR 56(2):77–84

Reif B, Hennig M, Griesinger C (1997) Direct measurement of angles between bond vectors in high-resolution NMR. Science 276(5316):1230–1233

Rennella E, Cutuil T, Schanda P, Ayala I, Forge V, Brutscher B (2012) Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate. J Am Chem Soc 134(19):8066–8069

Riek R, Wider G, Pervushin K, Wüthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. P Natl Acad Sci USA 96(9):4918–4923

Robin M, Delsuc MA, Guittet E, Lallemand JY (1991) Optimized acquisition and processing schemes in three-dimensional NMR spectroscopy. J Magn Reson (1969) 92(3):645 – 650

Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol 263(5):627 – 636

Rovnyak D, Frueh DP, Sastry M, Sun ZYJ, Stern AS, Hoch JC, Wagner G (2004a) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. J Magn Reson 170(1):15 – 21

Rovnyak D, Hoch JC, Stern AS, Wagner G (2004b) Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. J Biomol NMR 30:1–10

Rovnyak D, Sarcone M, Jiang Z (2011) Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. Magn Reson Chem 49(8):483–491

Rule GS, Kevin HT (2006) Fundamentals of Protein NMR Spectroscopy. Springer, The Netherlands

Ruschak A, Velyvis A, Kay L (2010) A simple strategy for ^{13}C,^{1}H labeling at the Ile-γ_2 methyl position in highly deuterated proteins. J Biomol NMR 48(3):129–135

Sattler M, Maurer M, Schleucher J, Griesinger C (1995) A simultaneous ^{15}N,^{1}H- and ^{13}C,^{1}H-HSQC with sensitivity enhancement and a heteronuclear gradient echo. J Biomol NMR 5(1):97–102

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Mag Res Sp 34(2):93–158

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127(22):8014–8015

Schanda P, Kupče r, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33(4):199–211

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128(28):9042–9043

Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. P Natl Acad Sci USA 104(27):11,257–11,262

Schmieder P, Stern A, Wagner G, Hoch J (1993) Application of nonlinear sampling schemes to COSY-type spectra. J Biomol NMR 3(5):569–576

Schmieder P, Stern A, Wagner G, Hoch J (1994) Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain. J Biomol NMR 4(4):483–490

Schmieder P, Stern A, Wagner G, Hoch J (1997) Quantification of maximum-entropy spectrum reconstructions.

J Magn Reson 125:332–339

Shaka AJ, Keeler J, Freeman R (1984) Separation of chemical-shifts and spin coupling in proton NMR - elimination of dispersion signals from two-dimensional spectra. J Magn Reson 56(2):294–313

Shang Z, Swapna GVT, Rios CB, Montelione GT (1997) Sensitivity enhancement of triple-resonance protein NMR spectra by proton evolution of multiple-quantum coherences using a simultaneous ^{1}H and ^{13}C constant-time evolution period. J Am Chem Soc 119(39):9274–9278

Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656

Shapira B, Karton A, Aronzon D, Frydman L (2004) Real-time 2D NMR identification of analytes undergoing continuous chromatographic separation. J Am Chem Soc 126(4):1262–1265

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223

Shimba N, Stern A, Craik C, Hoch J, Doetsch V (2003) Elimination of $^{13}C^{\alpha}$ splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. J Am Chem Soc 125:2382–2383

Shore JE, Johnson RW (1980) Axiomatic derivation of the principle of maximum-entropy and the principle of minimum cross-entropy. IEEE T Inform Theory 26(1):26–37

Shrot Y, Frydman L (2003) Single-scan NMR spectroscopy at arbitrary dimensions. J Am Chem Soc 125(37):11,385–11,396

Shrot Y, Frydman L (2011) Compressed sensing and the reconstruction of ultrafast 2D NMR data: Principles and biomolecular applications. J Magn Reson 209(2):352 – 358

Sibisi S (1983) Two-dimensional reconstructions from one-dimensional data by maximum entropy. Nature (London) 301:134–136

Sibisi S, Skilling J, Breneton R, Laue E, Staunton J (1984) Maximum entropy signal processing in practical NMR spectroscopy. Nature 311:446–447

Sinha K, Jen-Jacobson L, Rule GS (2011) Specific labeling of threonine methyl groups for NMR studies of protein–nucleic acid complexes. Biochemistry-US 50(47):10,189–10,191

Skilling J (1984) Data-analysis - the maximum-entropy method. Nature 309(5971):748–749

Skilling J, Bryan RK (1984) Maximum entropy image reconstruction: general algorithm. Monthly Notices of the Royal Astronomical Society 211(1):111–124

Solomon I (1955) Relaxation processes in a system of two spins. Phys Rev 99:559–565

Sørensen M, Meissner A, Sørensen O (1997) Spin-state-selective coherence transfer via intermediate states of two-spin coherence in IS spin systems: Application to E.COSY-type measurement of J coupling constants. J Biomol NMR 10(2):181–186

Stanek J, Koźmiński W (2010) Iterative algorithm of discrete Fourier transform for processing randomly sampled NMR data sets. J Biomol NMR 47(1):65–77

Stanek J, Augustyniak R, Koźminski W (2012) Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. J Magn Reson 214:91–102

Stanek J, Nowakowski M, Saxena S, Ruszczyńska-Bartnik K, Ejchart A, Koźminski W (2013a) Selective diagonal-free ^{13}C,^{13}C-edited aliphatic-aromatic NOESY experiment with non-uniform sampling. J Biomol NMR 56(3):217–226

Stanek J, Podbevšek P, Koźminski W, Plavec J, Cevec M (2013b) 4D non-uniformly sampled C,C-NOESY experiment for sequential assignment of ^{13}C,^{15}N-labeled RNAs. J Biomol NMR 57(1):1–9

Stanek J, Saxena S, Geist L, Konrat R, Koźminski W (2013c) Probing local backbone geometries in intrinsically disordered proteins by cross-correlated NMR relaxation. Angew Chem Int Edit 52(17):4604–4606

Stern AS, Li KB, Hoch JC (2002) Modern spectrum analysis in multidimensional NMR spectroscopy: Comparison of linear-prediction extrapolation and maximum-entropy reconstruction. J Am Chem Soc 124(9):1982–1993

Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l_1-norm reconstruction. J Magn Reson 188(2):295–300

Sun ZY, Hyberts S, Rovnyak D, Park S, Stern A, Hoch J, Wagner G (2005) High resolution aliphatic side-chain assignments in 3D HCcoNH experiments with joint H–C evolution and non-uniform sampling. J Biomol NMR 32:55–60

Sundararajan D (2001) The Discrete Fourier Transform, Theory, Algorithms and Applications. World Scientific, Singapore

Swapna G, Rios C, Shang Z, Montelione G (1997) Application of multiple-quantum line narrowing with simultaneous ^{1}H and ^{13}C constant-time scalar-coupling evolution in PFG-HACANH and PFG-HACA(CO)NH triple-resonance experiments. J Biomol NMR 9(1):105–111

Tarczynski A, Allay N (2004) Spectral analysis of randomly sampled signals: suppression of aliasing and sampler jitter. IEEE T Signal Proces 52(12):3324–3334

Tikhonov A, Samarskij AA (1990) Equations of Mathematical Physics. Dover Publications, New York

Tikole S, Jaravine V, Orekhov VY, Güntert P (2013) Effects of NMR spectral resolution on protein structure calculation. PLoS ONE 8(7):e68,567

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278(5340):1111–1114

Tucker LR (1966) Some mathematical notes on 3-mode factor analysis. Psychometrika 31(3):279–279

Tugarinov V, Kay L (2004) An isotope labeling strategy for methyl TROSY spectroscopy. J Biomol NMR 28(2):165–172

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125(34):10,420–10,428

Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005a) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. P Natl Acad Sci USA 102(3):622–627

Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY (2005b) High-resolution four-dimensional ^{15}N-^{13}C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J Am Chem Soc 127(8):2767–2775

Velyvis A, Ruschak AM, Kay LE (2012) An economical method for production of ^{2}H,^{13}CH$_3$-threonine for solution NMR studies of large protein complexes: Application to the 670 kDa proteasome. PLoS ONE 7(9):e43,725

Venters R, Huang CC, Farmer I BennettT, Trolard R, Spicer L, Fierke C (1995) High-level ^{2}H/^{13}C/^{15}N labeling of proteins for NMR studies. J Biomol NMR 5(4):339–344

Vögeli B, Yao L (2009) Correlated dynamics between protein HN and HC bonds observed by NMR cross relaxation. J Am Chem Soc 131(10):3668–3678

Vuister GW, Clore GM, Gronenborn AM, Powers R, Garrett DS, Tschudin R, Bax A (1993) Increased resolution and improved spectral quality in four-dimensional ^{13}C/^{13}C-separated HMQC-NOESY-HMQC spectra using pulsed-field gradients. J Magn Reson Ser B 101(2):210–213

Weigelt J (1998) Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. J Am Chem Soc 120(41):10,778–10,779

Wen J, Wu J, Zhou P (2011) Sparsely sampled high-resolution 4-D experiments for efficient backbone resonance assignment of disordered proteins. J Magn Reson 209(1):94 – 100

Wen J, Zhou P, Wu JH (2012) Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl-methyl NOESY for large proteins. J Magn Reson 218:128–132

Werner-Allen JW, Coggins BE, Zhou P (2010) Fast acquisition of high resolution 4-D amide-amide NOESY with diagonal suppression, sparse sampling and FFT-CLEAN. J Magn Reson 204(1):173–8

Wirmer J, Schwalbe H (2002) Angular dependence of ^{1}J(N i,C$^{\alpha}i$) and ^{2}J(N i,C$^{\alpha}$(i - 1)) coupling constants measured in J-modulated HSQCs. J Biomol NMR 23(1):47–55

Wishart D, Sykes B (1994) The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4(2):171–180

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry-US 31(6):1647–1651

Wong L, Masse J, Jaravine V, Orekhov V, Pervushin K (2008) Automatic assignment of protein backbone resonances by direct spectrum inspection in targeted acquisition of NMR data. J Biomol NMR 42(2):77–86

Wüthrich K (1986) NMR of Proteins and Nucleic Acids. John Wiley and Sons, New York, NY

Wu JH, Fan JS, Pascal SM, Yang DW (2004) General method for suppression of diagonal peaks in

heteronuclear-edited NOESY spectroscopy. J Am Chem Soc 126(46):15,018–15,019

Xia Y, Man D, Zhu G (2001) 3D Haro-NOESY-CH_3NH and Caro-NOESY-CH_3NH experiments for double labeled proteins. J Biomol NMR 19(4):355–360

Xia Y, Yee A, Arrowsmith C, Gao X (2003) ^1H_C and ^1H_N total NOE correlations in a single 3D NMR experiment. ^{15}N and ^{13}C time-sharing in t_1 and t_2 dimensions for simultaneous data acquisition. J Biomol NMR 27(3):193–203

Xu Y, Lin Z, Ho C, Yang D (2005) A general strategy for the assignment of aliphatic side-chain resonances of uniformly ^{13}C,^{15}N-labeled large proteins. J Am Chem Soc 127(34):11,920–11,921

Xu YQ, Zheng Y, Fan JS, Yang DW (2006) A new strategy for structure determination of large proteins in solution without deuteration. Nat Methods 3(11):931–937

Yamazaki T, Yoshida M, Nagayama K (1993) Complete assignments of magnetic resonances of Ribonuclease-H from Escherichia-Coli by double-resonance and triple-resonance 2D and 3D NMR spectroscopies. Biochemistry-US 32(21):5656–5669

Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994) A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^2H labeled proteins with high sensitivity. J Am Chem Soc 116(26):11,655–11,666

Yang DW, Konrat R, Kay LE (1997) A multidimensional NMR experiment for measurement of the protein dihedral angle ψ based on cross-correlated relaxation between $^1H^\alpha$-$^{13}C^\alpha$ dipolar and ^{13}C' (carbonyl) chemical shift anisotropy mechanisms. J Am Chem Soc 119(49):11,938–11,940

Yang DW, Gardner KH, Kay LE (1998) A sensitive pulse scheme for measuring the backbone dihedral angle ψ based on cross-correlation between $^{13}C^\alpha$-$^1H^\alpha$ dipolar and carbonyl chemical shift anisotropy relaxation interactions. J Biomol NMR 11(2):213–220

Yang DW, Zheng Y, Liu DJ, Wyss DF (2004) Sequence-specific assignments of methyl groups in high-molecular weight proteins. J Am Chem Soc 126(12):3710–3711

Zawadzka-Kazimierczuk A, Kazimierczuk K, Koźmiński W (2010) A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins. J Magn Reson 202(1):109–116

Zawadzka-Kazimierczuk A, Kozminski W, Sanderova H, Krasny L (2012) High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52(4):329–337

Zhang O, Kay L, Olivier J, Forman-Kay J (1994) Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR 4(6):845–858

Zheng Y, Giovannelli JL, Ho NT, Ho C, Yang DW (2004) Side-chain assignments of methyl-containing residues in a uniformly ^{13}C-labeled hemoglobin in the carbonmonoxy form. J Biomol NMR 30(4):423–429

Zhou D, Shea J, Nieuwkoop A, Franks W, Wylie B, Mullen C, Sandoz D, Rienstra C (2007a) Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. Angew Chem Int Edit 119(44):8532–8535

Zhou DH, Shah G, Cormos M, Mullen C, Sandoz D, Rienstra CM (2007b) Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning. J Am Chem Soc 129(38):11,791–11,801

Zhu G, Yao X (2008) TROSY-based NMR experiments for NMR studies of large biomolecules. Prog Nucl Mag Res Sp 52(1):49 – 68